CROSS-LINGUISTIC PERCEPTION AND LEARNING OF MANDARIN CHINESE SOUNDS BY JAPANESE ADULT LEARNERS

by

PEIPEI WEI

A DISSERTATION

Presented to the Department of East Asian Languages and Literatures
and the Graduate School of the University of Oregon
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

December 2016

DISSERTATION APPROVAL PAGE

Student: Peipei Wei

Title: Cross-Linguistic Perception and Learning of Mandarin Chinese Sounds by Japanese Adult Learners

This dissertation has been accepted and approved in partial fulfillment of the requirements for the Doctor of Philosophy degree in the Department of East Asian Languages and Literatures by:

Kaori Idemaru	Chairperson
Lucien Brown	Core Member
Zhuo Jing-Schmidt	Core Member
Melissa Baese-Berk	Institutional Representative

and

Scott L. Pratt	Dean of the Graduate School

Original approval signatures are on file with the University of Oregon Graduate School.

Degree awarded December 2016

DISSERTATION ABSTRACT

Peipei Wei

Doctor of Philosophy

Department of East Asian Languages and Literatures

December 2016

Title: Cross-Linguistic Perception and Learning of Mandarin Chinese Sounds by Japanese Adult Learners

This dissertation presents a cross-linguistic investigation of how nonnative sounds are perceived by second language (L2) learners in terms of their first language (L1) categories for an understudies language pair---Japanese and Mandarin Chinese. Category mapping experiment empirically measured the perceived phonetic distances between Chinese sounds and their most resembling Japanese categories, which generated testable predictions on discriminability of Chinese sound contrasts according to Perception Assimilation Model (PAM). Category discrimination experiment obtained data concerning L2 learners' actual performance on discrimination Chinese sounds. The discrepancy between PAM's predictions and actual performances revealed that PAM cannot be applied to L2 perceptual learning. It was suggested that the discriminability of L2 sound contrasts was not only determined by perceived phonetic distances but probably involved other factors, such as the distinctiveness of certain phonetic features, e.g. aspiration and retroflexion.

The training experiment assessed the improvement of L2 learners' performance in identifying Chinese sound contrasts with exposure to high variability stimuli and feedback. The results not only proved the effectiveness of training in shaping L2 learners'

perception but showed that the training effects were generalizable to new tokens spoken by unfamiliar talkers.

In addition to perception, the production of Chinese sounds by Japanese learners was also examined from the phonetic perspective in terms of perceived foreign accentedness. Regression of L2 learners' and native speakers foreign accentedness ratings against acoustic measurements of their speech production revealed that although both segmental and suprasegmental variables contributed to the perception of foreign accent, suprasegmental variables such as total and intonation patterns were the most influential factor in predicting perceived foreign accent.

To conclude, PAM failed to accurately predict learning difficulties of nonnative sounds faced by L2 learners solely based on perceived phonetic distances. As Speech Learning Model (SLM) hypothesizes, production was found to be driven by perception, since equivalence classification of L2 sounds to L1 categories prevented the establishment of a new phonological category, thus further resulted in divergence in L2 production. Although production was hypothesized to eventually resemble perception, asynchrony between production and perception was observed due to different mechanisms involved.

CURRICULUM VITAE

NAME OF AUTHOR: Peipei Wei

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene, United States
Beijing Language and Culture University, Beijing, China

DEGREES AWARDED:

Doctor of Philosophy, East Asian Languages and Literatures, 2016, University of Oregon
Master of Arts, East Asian Languages and Literatures, 2011, University of Oregon
Bachelor of Arts, Japanese Language and Literature, 2005, Beijing Language and Culture University

AREAS OF SPECIAL INTEREST:

Phonetics
Second Language Acquisition
Speech Perception and Production
Foreign Accent

PROFESSIONAL EXPERIENCE:

Teaching Assistant, University of Oregon, 2008-2016

Technical Management, Panasonic R&D Co. Ltd, Beijing, China, 2005-2007

GRANTS, AWARDS, AND HONORS:

Siegel Graduate Travel Endowment Fund Scholarship, University of Oregon, 2012-2014

General University Scholarship, University of Oregon, 2010-2012

PUBLICATIONS:

Idemaru, K., & Wei, P. (2015). Strong Influence of Prosody on the Perception of Foreign Accent. *Proceedings of 18th International Congress of Phonetics*, Scotland, UK, August 10-14, 2015

Wei, P. & Idemaru, K. (2013). Acoustic Analysis of Perceived Accentedness in Mandarin Speakers' Second Language production of Japanese. Proceedings of Meetings on Acoustics (Vol. 19, p. 060089).

ACKNOWLEDGMENTS

I wish to express sincere appreciation to my dissertation advisor Professors Kaori Idemaru for her advice and assistance during the whole dissertation writing process. I will not be able to complete this project without her help. In addition, special thanks are due to all my committee members: Professor Lucien Brown, Professor Zhuo Jing-Schmidt, and Professor Melissa Baese-Berk, who had provided insightful comments for my dissertation. I also thank all the Japanese and Chinese participants for their time and participation in all the experiments. I am thankful to my family, colleagues, and friends who supported me through this long journey. I also need to thank the department of East Asian Languages and Literatures at University of Oregon for awarding me with Siegel Graduate Travel Endowment Fund Scholarship, which provided me with indispensable financial support so that I could fly to China to collect precious data that is essential for the completion of this project.

TABLE OF CONTENTS

Chapter Page

Chapter Page

Chapter Page

LIST OF FIGURES

Figure Page

LIST OF TABLES

Table Page

Table Page

Chapter I

INTRODUCTION

Human beings are endowed with the ability to perceive a wide range of speech sounds at birth, either native or nonnative, but this ability gradually attenuates with increased exposure to a specific language environment. Infants begin to develop language-specific perceptual system and start to lose the language-universal perception by attending more to sound contrasts that have distinctive meanings in their native languages and at the same time paying less attention to sounds that do not distinguish meanings (Aslin & Pisoni, 1980; Kuhl, Williams, Lacerda, Stevens, & Lindblom, 2006; Tsao, Liu, & Kuhl, 2006; Werker, 1989). The developmental changes of perceptual system had been extensively researched in infant perception studies, and were found to usually occur at the first year of life (Werker & Tees, 1984a). During this very first stage of language learning, infants' perception of nonnative phonetic contrasts deteriorates rapidly. For example, English infants started to lose their ability to distinguish Hindi and Salish consonant contrasts at 10-12 months year old (Werker, Gilbert, Humphrey, & Tees, 1981; Werker & Tees, 1983, 1984a). The same degeneration was found not only in segmental domain regarding perceptional abilities of vowels and consonants but in suprasegmental domain. It was found English infants' performance on discrimination of speech (lexical) tones deteriorated between 6 and 9 months of age in comparison to Chinese infants (Mattock & Burnham, 2006). At the same time, this attenuation with increased experiences in a specific language was found to be compensated by increased ability of perceiving phonemic contrasts in their native languages. In a study comparing American and Japanese infants' performances on [r] vs. [l] contrast, both groups

performed equally well initially. By 10-12 months of age, American infants' performance improved significantly, while a decline was observed in Japanese infants' performance (Kuhl, et al., 2006). Similar results were also reported in Tsao et al (2006)'s cross-linguistic study on discrimination of affricate-fricative contrasts in English and Mandarin by infants. Whether tested with English or Mandarin affricate-fricative contrasts, both English- and Mandarin-learning infants did not differ at the age of 6-8 months but diverged at 10-12 months. Both English and Mandarin infants performed better in distinguishing affricate-fricative contrasts in their native languages while worse on contrasts in nonnative languages.

This process of perceptual reorganization at early infancy resulted in enhanced ability to perceive native phonemic contrasts and reduced sensitivity to nonnative contrasts, which on the other hand also posed great difficulty for adult second language (L2) learners to perceive and produce non-native sounds. To acquire a second language, learners have to learn how to make distinctions on the sounds in the target language as what they do in their native languages. However, adult learners differ from infants in the way that they have their native phonological system in place, which serves as a "sieve", filtering out phonetic signals that are not accommodated by this system (Trubetzkoy, 1969). It has been well documented that the performance of adult learners on discriminating nonnative sounds was considerably worse than that of native speakers (Goto, 1971; Miyawaki et al., 1975; Sheldon & Strange, 1982; Tees & Werker, 1984; Trehub, 1976; Werker, et al., 1981; Werker & Logan, 1985). This performance difference was found to be due to attentional or cognitive changes rather than sensorineural loss (Werker & Tees, 1984b). Therefore, the acquisition of a second language is a process of

combating first language (L1) bias by modifying the speech perceptual system that has already been shaped by life-long exposure to the native language.

This dissertation aims to contribute to the body of cross-linguistic studies by investigating how nonnative sounds are perceived by L2 learners in terms of their L1 "phonological filter" for a understudied language pair---Japanese-Mandarin Chinese and how the perception is related to difficulties they encounter when trying to learn to perceive and produce certain nonnative sounds. The influences of L1 phonology on acquisition of L2 were well studied and had given birth to a number of theoretical models, such as Contrastive Analysis Hypothesis (CAH) (Lado, 1957) and Differential Markedness Hypothesis (DMH) (Eckman, 1977) in attempt to account for the difficulties L2 learners face when trying to acquire nonnative sounds.

1.1. Japanese L2 Learners' Perception and Production of Chinese Sounds

Out of all the theoretical models, two most recent theoretical models are Flege's SLM (Speech Learning Model) (James Emil Flege, 1991; James E Flege, 1995) and Best et al's PAM (Perceptual Assimilation Model) (Best, 1995). SLM and PAM are constructed based on the interaction between phonetic space of native phones and phonetic proximity between native and nonnative sounds to predict the degree of difficulty of L2 perception and production. Both models postulate that L2 learners would rely on their L1 sound system in L2 perception and production at least at the onset of L2 learning. The fundamental claim of PAM is that discrimination of a nonnative contrastive pair (e.g., English [r] vs. [l] for Japanese L2 learners) depends on the similarities (or dissimilarities) between each nonnative sound and its most resembling L1 sound. Whether or not the nonnative sound pairs are perceived as a “good,” “acceptable” or

“deviant” exemplars of the L1 phonetic category predicts the perceptual discriminability of the L2 sound pair. For example, English [r] vs. [l] may be perceived as equally “deviant” exemplars of a single Japanese category---tap [ɾ], thus the discrimination between these two sounds is predicted to be "difficult" by PAM, which is supported by findings in previous studies. SLM, on the other hand, is built upon studies on experienced L2 learners across different languages. It focuses on making predictions on changes in L2 learners’ long-term learning experiences. SLM hypothesizes that the more dissimilar an L2 sound is perceived in comparison to its closest L1 sound, the easier this L2 sound can be discerned and learned. On the other hand, a new category will not be established for an L2 sound if it is classified as equivalent to an existing L1 phonetic category, and thus the production of this L2 sound will eventually resemble this L1 category. Both models mainly focus on segmental (vowels and consonants) learning and did not take suprasegmentals into consideration. The detailed hypotheses of both models will be laid out in Chapter 2. Since the primary objective of this dissertation is to examine the perception and production of Chinese sounds by Japanese L2 learners, these two models serve as theoretical frameworks that drive the discussion of findings in this dissertation to the theoretical level in order to contribute to the general knowledge of SLA speech perception and learning. Since both models regard the perception and production of L2 sounds as highly related to L1 sound categories, I will provide a detailed comparison of Chinese and Japanese phonological system in the following section.

1.2. A Comparison of Chinese and Japanese phonological system

This section will provide a comprehensive comparison of Japanese and Chinese phonological system in terms of segmeantals and suprasegmenals in order conceptualize

the tasks Japanese L2 learners face when trying to learn Mandarin Chinese. Mandarin Chinese have a stock of five monophthong vowel phonemes: [a, i, y, ə, u]. According to articulatory features of degree of openness or tongue height, these five vowels can be classified into high [i, u, y], mid ([ə]) and low ([a]) vowels. In terms of place of articulation, they can be classified into front ([i, y]), central ([ə]), and back ([u, a]) vowels. In terms of lip position, they can be divided into unrounded ([a, i, ə]) and rounded ([y, u]) vowels. Some of these vowels also have allophones that appear as variations in particular context (Lin, 2007). Vowels [æ, ɑ, ɛ] are allophones of [a] and [e, o, ɤ] are allophones of [ə]. Also, there are three glides [j, w, ɥ] in Chinese which occur in syllable onset position as allophones of three high vowels respectively ([i, u, y]) (Lin, 2007). Besides these monophthongs, there are four diphthongs comprised by two monophthongs: [ai, ɑu, ei, ou]. In addition, there are also retroflex vowels in Chinese, the pronunciations of which require a curled backed tongue to be raised to the post-alveolar region. The status of retroflex vowels is still uncertain, since scholars still have debates on whether it is a single vowel, a syllabic consonant, a diphthong, or vowel plus a consonant (Lin, 2007).

Standard (Tokyo) Japanese have five short vowel phonemes: [a, i, ɯ, e, o]. According to the articulatory features of degree of openness or tongue height, they can be classified into high ([i, ɯ]), mid ([e, o]) and low ([a]) vowels. In terms of place of articulation, they can be classified into front ([i, e]), central ([a]), and back ([o, ɯ]) vowels (Labrune, 2012). In terms of lip position, all vowels are unrounded ([a, i, ɯ, e]) except [o]. Each of these short vowels has their own counterpart of long vowels, which are usually transcribed as [a:, i:, u:, e:, o:]. The difference between short and long vowel

entirely lies in duration (Vance, 2008). Although there are also durational differences between diphthongs and monophthongs in Chinese (Svantesson, 1984), all diphthongs in Chinese consist of different vowels, and the length of monophthongs is not phonemic. Similar to Chinese, there are also two semivowels or glides in Japanese, which occur in syllable initial position [j, ɥ] as the allophones of high front vowel [i] and high back vowel [ɯ]. As we compare Chinese vowel inventory with Japanese counterpart, we can see that Chinese have two vowels that Japanese lacks: the rounded front vowel [y] and central mid vowel [ə]. Also, Japanese does not have diphthongs and retroflex vowels. Although both Chinese [u] and Japanese [ɯ] are characterized as back front vowel, it is also noted that it is actually more appropriate to label Japanese [ɯ] as "central" (Vance, 2008). Articulatorily, the articulation of Chinese [u] involves lip intrusion while the pronunciation of Japanese [ɯ] does not, instead lip compression might occur in some speech (Lin, 2007; Vance,2008)

Chinese consonants can be divided into bilabial ([p, p^h, m, w, ɥ]), labio-dental ([f]), dental ([t, t^h, s, ts, ts^h, n, l]), post-alveolar ([ʂ, tʂ, $tʂ^h$, ɹ]), alveolo-palatal ([ɕ, tɕ, $tɕ^h$]), palatal ([j, ɥ]) and velar ([k, k^h, x, w, ŋ]) consonants based on different place of articulation. [t, t^h, n, l] can also be alveolar depending on different speakers (Lin, 2007). Based on the manner of articulation, these consonants can be classified into stop ([p, p^h, t, t^h, k, k^h]), fricative ([f, s, ʂ, ɕ, x]), affricate ([ts, ts^h, tʂ, $tʂ^h$, tɕ, $tɕ^h$]), nasal([m, n, ŋ]), approximant([ɥ, w, ɹ, j]), lateral ([l]). As mentioned above, three glides [j, w, ɥ] are allophones of three high vowels ([i, u, y]) respectively which occur in syllable onset position (Lin, 2007). Alveolo-palatal ([ɕ, tɕ, $tɕ^h$] are sometimes considered to be

allophones of alveolar ([s, ts, ts^h]) since these alveolar consonants usually undergo palatalization when preceding high vowels or glides.

Based on different place of articulation, Japanese consonants can be divided into bilabials ([p, b, m, ɸ, ß]), alveolar ([t, d, s, z, ts, n, ɾ]), palatals ([ɕ, ʑ, tɕ, y]), velar ([k, g, w, ŋ]) and glottal ([h]). Based on manner of articulation, these consonants can be classified as stop ([p, b, t, d, k, g]), fricative ([ɸ, ß, s, z, ɕ, ʑ, h]), affricate ([tɕ, ts]), nasal ([m,n,ŋ]), glide ([w, y]), and tap/flap ([ɾ]). Among the fricatives, [ɸ] and [ç] are the allophones of [h], although [ɸ] has arguably obtained independent phonemic status in Japanese, since it can occur with any vowels in loanwords (Vance, 2008). Voiced fricatives [ß, ð and ɣ] occur in Japanese in rapid speech as allophones of [b, d, g] respectively. [ŋ] is not a phoneme in Japanese. Some instances of [ŋ] is an allophone of [n] when it occurs before a velar sound while some instances of [ŋ] functions as an allophone of [g] (Tsujimura, 2013). Some instances of [m] functions as allophone of [n] when it occurs before bilabial and other instances of [m] functions as phoneme [m] (Tsujimura, 2013). Generally [ts, tɕ, ɕ, ʑ, ɸ, ç]) are considered to be allophones of [t, t, s, z, h, h] respectively, although this is still controversial because some of them could also occur independently as contrastive phonemes (Tsujimura, 2013). In addition to all these singleton consonants, Japanese also have geminate consonants, which do not exist in Chinese consonant inventory. Chinese stops are distinguished by aspiration while Japanese stops are distinguished by voicing. Actually, all Chinese stops are voiceless. Acoustically, aspirated consonants have much longer VOT (voice onset time) than unaspirated consonants. Although aspiration does occur in Japanese voiceless stops, this feature is not phonemic and thus does not distinguish meanings. Japanese stops [p, t, k]

are usually produced as aspirated in word-initial position or in accented syllable and unaspirated elsewhere (Vance, 2008). Acoustically, Japanese voiceless stops have VOTs in between those prototypical voiceless unaspirated stops and voiceless aspirated stops. It is also claimed that VOTs of Japanese voiceless stops fall between the two general groupings of voiceless stops: short lag (0-25ms for unaspirated stops) and long lag (60-100ms for aspirated stops) in world's languages (Riney, Takagi, Ota, & Uchida, 2007). Similar to stops, Chinese fricative and affricates are only distinguished by aspiration instead of voicing. Also, we can see that Japanese does not distinguish lateral and tap phonemically. It only has a tap phoneme [ɾ], which a diverse variation of realization:[ɹ, ɻ, ɖ, ɭ, r, l] (Best & Strange, 1992; Ingram & Park, 1998; Okada, 1991; Smith & Kochetov, 2009). In contrast, Chinese has both lateral [l] and retroflex approximant [ɹ] as phonemes (Lee & Zee, 2003). Instead of having labiodentals phoneme [f] as in Chinese, only bilabial fricative [ɸ] occurs in Japanese as an allophone of glottal fricative [h]. Compared to Chinese, Japanese lacks the category of post-alveolar consonants which is also the so called retroflex.

As for prosodic system, Chinese is characterized as a tonal language which employs the pitch height to distinguish meanings. There are basically four phonemic types of pitch patterns in Chinese: high level tone (55), high rising tone (35), low falling-rising tone (214) and high falling tone (51) with "5" indicating the highest pitch and "1" indicating the lowest pitch height (Lin, 2007). In addition, there is a neutral tone, the pitch value of which is determined by the preceding tone. The phenomenon of tone sandhi happens when a sequence of tones occur in certain combinations or tonal environments. Since both tone and intonation make uses of pitch variation, to avoid

conflict of these two, Chinese usually applies final particles which are initially neutral tones as the primary medium for realization of intonation (Lin, 2007). As for rhythm, Chinese is characterized as syllable-timed language, in which syllable functions as the isochronous unit of speech (Mok, 2009).

In contrast, Japanese is characterized as pitch accent language. The acoustic correlate of Japanese pitch accent is a F0 fall. Tones are lexical property in Chinese, and Japanese pitch accents are also lexical, which means that every lexical item has its predetermined pitch pattern. A lexical word can have either accent or no accent. Although Chinese tone is assigned for each individual syllable, the notion of pitch accent in Japanese only comes into play in the domain of word or phrase. Based on certain phonological rules, the entire pitch pattern of a word or phrase can be determined by simply identifying the location of the pitch accent. This pitch pattern also can undergo change when words are combined together into a new compound word. As for accent pattern of an unaccented phrase, there is usually a relatively low pitch at the beginning, a relatively succeeding high pitch, and a relatively low pitch at the end (Venditti, 2006). For a phrase with accent, there is an additional steep pitch fall which characterizes the accent, which is called accentual peak (Vance, 2008). In contrast to Chinese being syllable-timed, Japanese is Japanese is characterized as a mora-timed language which mora is considered as the isochronous unit of utterance.

1.3 Japanese L2 learners' perception and production of Chinese sounds

The above comparison provides us with the basis on which we can discuss the learning difficulties of Mandarin sounds for Japanese-speaking learners (referred as Japanese L2 learners in this dissertation) with reference to PAM and SLM. Since PAM

grounds its hypotheses with regard to discriminability of nonnative sounds based on their perceived similarities or dissimilarities to the closest L1 categories, although we can make speculations based on comparison of Japanese and Chinese phonology above, it is hard to make concrete predictions. Only a few studies attempted to evaluate PAM hypotheses by empirically measuring the perceived phonetic differences between L2 sounds and the most resembling L1 categories, but most of them suffer from limitations either in terms of scope or methodological issues (Best, McRoberts, & Goodell, 2001; Guion, Flege, Akahane-Yamada, & Pruitt, 2000; Harnsberger, 2001; Schmidt, 2007; R. P. Wayland, 2007). Chapter 1 aims to contribute to the research efforts of empirical evaluation of PAM framework by conducting cross-linguistic investigation of L2 perception for the language pair of Chinese and Japanese. Specifically, Chapter 1 first presents a categorical mapping study which empirically measures the perceived phonetic distances between a comprehensive list of Chinese vowels and consonants and their most resembling Japanese categories. This data enables us to make concrete predictions on discriminability of certain sound contrasts based on five assimilation types defined by PAM (see Chapter 2 for details). In order to verify the predictions of PAM's prediction, we need actual data with regard to how Japanese L2 learners perform in discriminating Chinese sounds contrasts. Accordingly, a category discrimination experiment followed category mapping experiment was designed to obtain this data so as to enable the verification of PAM's predictions by comparing the predicted discriminabilities of a select 15 Chinese sound contrasts and the actual performances. It should be reminded that the target of PAM's predictions is the discriminability of unknown nonnative sounds while we are interested in exploring the learning difficulties of L2 sounds in the context

of Second Language Acquisition (SLA). Thus testing Japanese L2 learners with Chinese language experience on discriminating Chinese sound contrasts made it possible to testify whether PAM could be extended to SLA field. A discussion based on the comparison between PAM's predictions and discrimination results are detailed in Chapter 2.

Despite the great difficulties adult L2 learners may encounter when trying to learn nonnative sounds, the flexibility of human speech perceptual system has been extensively proved by listeners' improved performance after receiving laboratory-based auditory training (Bradlow, Akahane-Yamada, Pisoni, & Tohkura, 1999; Hirata, Whitehurst, & Cullings, 2007; Iverson, Hazan, & Bannister, 2005; Lively, Logan, & Pisoni, 1993; Logan, Lively, & Pisoni, 1991; McClaskey, Pisoni, & Carrell, 1983; Strange & Dittmann, 1984; Wang, Jongman, & Sereno, 2003; Wang, Spence, Jongman, & Sereno, 1999; Wong, 2012). Discriminabilities based on the results from categorical mapping experiment in Chapter 2 informed us that certain Chinese sound contrasts are harder to discriminate than others. Accordingly, we are not only interested in finding whether intensive laboratory training can facilitate Japanese L2 learners' perception of Chinese sounds, but interested in examining whether training effects, if any, are the same for "easy" and "hard" sound contrasts. Chapter 3 presents a training study which employed HVPT (High Variability Phonetic Training) paradigm that has been proved to be effective in modifying cross-linguistic perception of nonnative sounds (Hirata, et al., 2007; Iverson, et al., 2005; Logan, et al., 1991; Wang, et al., 2003; Wang, et al., 1999; Wong, 2012). One stop contrast with predicted "hard" discriminability and one affricate contrast with "easy" discriminability were selected as the two sound contrasts that Japanese L2 listeners were trained and tested on. The training followed a classic training

experiment design and included pretest, posttest and generalization tests (Logan, et al., 1991; Wang, et al., 2003; Hirata, et al., 2007). The training effects on these two consonant contrasts are discussed in Chapter 3.

However, the task of L2 learning is not accomplished by simply being able to distinguish nonnative sounds, the ultimate goal is speech production. If L2 learners can not only easily discriminate nonnative sounds but produce them in a native-like fashion, L2 learning would not be a linguistics field that has attracted so many research interests. Similar to poor performances on the discrimination of nonnative sounds, most L2 learners produce L2 speech with a noticeable foreign accent even after years of even lifelong language experiences. Foreign accent phenomenon have been extensively researched in the field of SLA. However, most of the research efforts have been devoted into exploring subject-related factors such as AOL (Age of Learning), LOR (Length of Residence) and others (Long, 1990; Patkowski, 1990; Piske, MacKay, & Flege, 2001; Scovel, 1988; Suter, 1976), while only a few studies attempted the investigation of foreign accent from a phonetic perspective (Missaglia, 1999; Munro, 1995; R. Wayland, 1997) . Chapter 4 first presents a production study investigating how Japanese L2 learners' speech production of Mandarin Chinese sounds differ from native speakers' by conducting acoustic measurements. A follow-up accentedness rating study was designed to find whether these acoustic measurements are perceptually linked to the perception of foreign accent by relating acoustic differences with foreign accentendness ratings together. As introduced above, different from PAM, SLM as a speech learning model considers perception to precede production and explicitly hypothesizes on the production of nonnative sounds based on how they are perceived by L2 learners. The acoustic

measurements of Japanese L2 learners' production of Mandarin Chinese sounds are presented and discussed with reference to SLM theoretical model in Chapter 4. The findings concerning the relationship between acoustic measurements and the perception of foreign accent are also summarized subsequently.

Chapter II

JAPANESE L2 LEARNER'S CATEGORY MAPPING AND DISCRIMINATION FOR MANDARIN CHINESE SOUNDS

2.1. Introduction

As introduced in Introduction chapter, two most influential phonetic models of L2 speech learning are PAM (Perceptual Assimilation Model) developed by Best and his colleges (Best et al, 1988; Best, 1993, 1995) and Flege's (1995) SLM (Speech Learning Model). PAM as a cross-linguistic speech perception model mainly focuses on the influence of native sound categories on the perception of novel nonnative sounds. The fundamental claim of PAM is that the perception of a L2 sound depends on the similarities or dissimilarities between this L2 sound and the most resembling L1 sound in native phonological system. There are three cases for L2 sound assimilation: (1) a L2 sound is assimilated to a L1 category as either a good exemplar, or an acceptable but not ideal exemplar, or an apparently deviant exemplar of that category; (2) a L2 sound is assimilated to non-L1 category when it is perceived to fall into L1 phonological space as speech sound yet not heard as a clear exemplar of any native categories, i.e., this L2 sound could be perceived to fall in between more than one L1 categories; (3) a L2 sound is not heard as speech sound because it is does not fall into the L1 phonological space. Based on these three assimilation patterns, predictions on discriminability between L2 sound pairs can be summarized to include the following six types: (1) *Two-Category Assimilation (TC Type):* if the L2 sound pairs are assimilated into two different L1 categories respectively, and the discrimination is predicted to be excellent; (2) *Category-*

Goodness Assimilation (CG Type): if both L2 sounds are assimilated to the same L1 category yet differ in degrees of proximity to that category, the discrimination between these two sounds is predicted to be moderate to very good, depending on the degree of resemblance to L1 categories for each sound. For instance, for two L2 sounds that are assimilated into the same L1 category as "deviant" and "good" exemplar respectively, the discriminatation is predicted to be very good. (3) *Single-Category Assimilation (SC Type)*: if both L2 sounds are assimilated into the same L1 category with equal degree of proximity to that category, discrimination is predicted to be poor. (4) *Both Uncategorizable (UU Type):* if both L2 sounds are perceived to fall into L1 phonological space yet not heard as a clear exemplar of any L1 categories, the discrimination is predicted to be poor to very good, depending on the similarities between these two L2 sounds as well as their proximity to L1 categories. (5) *Uncategorized versus Categorized (UC Type):* if one L2 sound is assimilated into a L1 category, while the other is heard to be a speech sound yet falls outside any L1 category, the discrimination is predicted to be very good. (6) *Nonassimilable (NA Type):* if both L2 sounds are perceived to be non-speech sounds, the discrimination is predicted to be good to very good.

In contrast to PAM, the theoretical framework of Flege (1995)'s SLM (Speech Learning Model) is built upon research findings about learning process of experienced L2 learners and mainly focuses on explaining age-related changes in speech perception and production of L2 vowels and consonants. There are a few assumptions that SLM grounds its claims upon. For instance, SLM regards that perception precedes and eventually shapes production. Also, under the framework of SLM, acquisition of L1 and L2 sound goes through the same process, thus the mechanism used in L1 acquisition can be applied

to L2 learning. L1 and L2 sounds share the same phonological space. Accordingly, phonetic categorization for L1 sounds evolve throughout the learners' lift span rather than stay stagnant once established, since they need to be adjusted in relation to all L1 and L2 sounds learned later on. Based on these assumptions, seven hypotheses are proposed in SLM. **H1:** Sounds in L1 and L2 are perceived in relation to one another at allophonic level rather than more abstract phonemic level and is position-sensitive. **H2:** A new phonetic category of a L2 sound can be established if the phonetic dissimilarities between this L2 sound and its closest L1 counterpart can be discerned. **H3:** The more dissimilar the L2 sound and preexisting L1 phonetic categories are perceived to be, the more likely the phonetic dissimilarities between the L2 and L1 sounds can be discerned. **H4:** As AOL (Age of Learning) increases, it is more unlikely for L2 learners to discern phonetic differences between L1 and L2 sounds that are not phonemic in their L1s. **H5:** Equivalence classification might prevent the establishment of a new phonetic category. Perpetually linked L1 and L2 sounds will be perceived as belonging to the same phonetic category which thus leads to undifferentiated sounds in production. **H6:** The phonetic category of a L2 sound established by L2 learners might be different from that of a monolingual. This is because L1 and L2 share the same phonological space, thus this L2 phonetic category might be established in the way such that it is distinctive to the preexisting L1 category. Also, it is also possible that this phonetic category is established based on different features or feature weights than a monolingual. **H7:** The production of a L2 sound eventually corresponds to properties of the phonetic category established for this sound.

Despite different focuses, these two speech learning models share the same notion

that the easiness or hardness to learn (no matter it is to discriminate or to produce) a L2 sound is dependent on the perceived similarities or dissimilarities between L2 and the closest L1 category this sound is categorized into. Accordingly, as long as we have access to this category similarities/dissimilarities information, we can make predictions on discrimination difficulties of certain nonnative sounds under the framework of these models and further empirically evaluate our predictions by obtaining actual discrimination scores. However, very limited research efforts have been devoted to providing such empirical data.

Best et al (2001) evaluated PAM predictions on discrimination difficulty order of three assimilation types (two-category (TC), category-goodness (CG) and single-category (SC)) by examining the perception of Zulu and Ethiopian Tigrinya consonant contrasts by native speakers of American English with no experience of the target languages. After completing the AXB discrimination tests in which listeners were asked to select the odd item when presented with three stimuli, listeners were then instructed to complete questionnaire tasks to write the most resembling English category and describe the way the stimuli sounded to them in English orthography. Accordingly, assimilation patterns of sound contrasts for discrimination tests were determined from listeners' English transliteration and descriptions of any consonantal differences between the target consonant and the English category they had classified. For instance, if Zulu/Tigrinya consonant contrasts were written down in the same English spelling with no description of consonantal differences, this contrast was determined to be TC assimilation pattern. The findings supported TC > CG > SC discrimination difficulty order predicted by PAM. Although this study provided important insight for future studies which aim to

empirically evaluate speech learning/perception theoretical framework, it was limited in its scope. Rather than presenting comprehensive categorical mapping data for each consonant directly, only assimilation patterns for certain consonant contrasts were of interest in this study. In addition, this study lacked a quantitative system to evaluate the degree of goodness-of-fit for each categorization.

Similar to previous study, Harnsberger (2001) also tested PAM's predictions on varying discriminability of five assimilation types: two-cateogry (TC), uncategorizable-categorizable (UC), both uncategorizable,(UU), category-goodness (CG), and single-category (SC). In this study, Malayalam, Marathi, and Oriya nasal consonants were presented to seven listener groups for forced-choice identification tests followed by AXB discrimination tests. Compared to Best et al's study (2001), one of the methodological improvement was that instead of relying on qualitative description, it employed a quantitative system to gauge category goodness on a 5-point scale. In identification tests, listeners were instructed to select a native category from a closed response sets. No specific justification was given regarding why forced-choice was used instead of asking listeners to freely write down the identified category in their L1. Probably it was due to these methodological differences, contrary to Best et al's (2001) findings, the results in this study indicated that not all PAM's predictions were born out. Specifically, the descending discriminability ordering of TC > CG > SC were not supported since they were found to be of the same discrimination difficulty level.

Guion et al (2000) investigated the perceived phonetic distances for a limited number of English consonants in terms of proximity to their classified Japanese categories and further explored the relationship between the phonetic distances and actual

discriminability of these sounds. In categorical mapping experiment, nine native Japanese speakers who know little English were asked to write down the closest Japanese category for each English consonant they were presented and provide goodness-of-fit rating on a 5-point scale. Different from previous studies in which category identification was analyzed in comparison with goodness rating to determine assimilation patterns, this study invented a metric called fit index so as to incorporate the identification percentage and goodness-of-fit rating of the identified Japanese category into one single measure by taking the product of these two. Based on fit index measures, each consonant was then divided into "good", "fair" or "poor" exemplar groups for the Japanese category it was classified into, which readily generated testable predictions on discriminability of a few consonant contrasts based on PAM framework. A categorical discrimination experiment was further conducted on three groups of Japanese L2 learners varying in English experience using AXB experiment design. Since PAM only focuses on perception of unknown non-native sounds and SLM mainly addresses the changes of L2 learning over life-span, this design made it possible to testify possible extension of PAM to L2 acquisition context and evaluate SLM framework in terms of learning effect for relatively inexperienced L2 learners. The discrimination results confirmed that the discriminability of consonant pairs was relevant to perceived phonetic distances represented by fit index. However, out of three English consonant contrasts, the discriminability for one contrast ([s-θ] UC Type) was not found as predicted by PAM, indicating the necessity of revision of PAM framework in order to account for predictions in L2 acquisition setting. Overall, this study introduced a few sophisticated research techniques such as using fit index to measure the perceived phonetic distances and calculating A prime scores to provide

unbiased metric of perceptual sensitivity, which served as very useful references for current study. The results also confirmed that language experience did play a role in discrimination performance on nonnative sounds since high experienced group outperformed less experienced group. However, similar to PAM, it was found that SLM cannot be extended to the early stage of L2 learning in general either, since only one out of the three consonant contrasts showed effect of learning.

Wayland (2007) was not only interested in examining the relationship between category identification and discrimination but whether the stimulus presentation contexts used in identification tests had any effects on discrimination performances. To do this, Wayland (2007) selected the language pair Thai---Korean and conducted two sets of identification and discrimination tests with Thai listeners identified and discriminated Korean stops and Korean listeners identified and discriminated Thai stops separately. In each set of experiment, two different identification tests, one presented "single" stimulus of in isolation and the other used "triadic" stimulus to present target stop as X of three sounds in the form of AXB, were administrated. Listeners were asked to identify each target stop in terms of their native consonant (Thai/Korean) category and rate the goodness-of-fit score on a scale from 1 to 5. In both experiments, AXB discrimination tests were administrated to obtain the actual discrimination scores. Predicted discrimination scores generated from identification tests with two different presentation contexts ("single" or "triadic") were correlated with actual discrimination scores. It was found that in some cases, the identification of target stops presented did differ depending on the stop consonants it was present in (i.e., what A or B is). In addition, actual discrimination score correlated better with identification data obtained from triadic than

single presentation contexts, especially for Thai listeners identifying and discriminating Korean stops. Although it only investigated category mapping on stops between the language pair of Thai and Korean, this study drew researchers' attention to the importance of matching the identification and discrimination tasks to avoid any undesired methodological effects on the results. We applied this message to our experiment design and used AX forced choice tests instead of AXB design which were widely used in previous category discrimination study. In AX forced choice tests, participants were asked to indicate whether the sound pairs they hear are the same or different. This matched the single presentation technique employed in the categorical mapping experiment presented in this chapter.

Schmidt (2007) also noticed the limited empirical categorical mapping data available for testing existing L2 speech perception models, and thus added research effort in this field by investigating how Korean consonants were perceptually mapped to English categories. 19 Korean syllable-initial consonants (stops [p, p^h, p*, t, t^h, t*, k, k^h, k*], nasals [m, n], affricates[s, s*], fricatives [tʃ, $tʃ^h$, tʃ*], glottal [h] and approximant [j]) were presented in three different vowel contexts ([i, a, u]) to 20 English native speakers with no experience of learning Korean. The listeners were asked to the identify the closest English category for the consonant they heard and rate the category goodness on a scale from 1 to 5. The results were presented with both identification accuracy in percentage and goodness-of-fit rating. It was also found that vowel contexts ([i, a, u]) following target consonants affected category identification along with its goodness ratings. Accordingly, in the experiments carried out in this chapter, all the Chinese consonant stimuli were produced in the same vowel context ([u]) and all the vowel

stimuli were produced in the same consonant context ([l]) so as to prevent the possible interference in perception by phonetic environment.

All the studies reviewed above only focused on a limited number of consonants for cross-linguistic perceptual mapping, this chapter however, aims to provide a comprehensive picture regarding how Mandarin Chinese sounds are perceived by Japanese native speakers in general by examining a full inventory of Chinese consonants and most Mandarin Chinese monophthong vowels and their allophones. Since the interest of this chapter mainly lies in the perception of Chinese sounds by Japanese L2 learners at the early stage of learning Mandarin Chinese, PAM instead of SLM is of relevance and used as the theoretical model to derive predictions and evaluation. As discussed above, PAM grounds its theoretical framework in the perception of nonnative sounds by listeners with no language learning experience. Accordingly, similar to Guion et al (2001)'s study, evaluating predictions on discriminability by PAM using discrimination data from Japanese L2 learners makes it possible to testify the possible extension of PAM to the context of L2 learning.

In this chapter, we report two perception experiments: Category mapping and discrimination experiments. Category mapping experiment aimed to contribute to the research efforts on cross-language perception in SLA by obtaining perceptual mapping data for one specific language pair---Chinese-Japanese. The primary objective of this experiment is to investigate how Japanese native speakers perceive different Mandarin Chinese sounds (both vowels and consonants) in terms of the categories in their native language by measuring the perceived phonetic proximity of Chinese sounds to Japanese categories. Predictions derived from PAM (Perception Assimilation Model) on

discriminability of L2 sound pairs are sometimes constructed based on phonological comparisons between L1 and L2. However, here, we attempt to empirically measure the perceived distance between L2 and L1 sounds by conducting Category mapping experiment to better understand the task of L2 learning the Japanese learners face in learning Chinese sounds. Based on results of this experiment, now can propose testable predictions on discrimination difficulties of sound pairs based on PAM**.** At last, we also need to obtain data regarding how Japanese L2 learners of Mandarin Chinese actually perform in the tasks of discriminating Chinese sound pairs with varying degrees of perceived phonetic distances to evaluate our predictions. We are interested in examining whether Japanese L2 learners had more difficulty in discriminating Chinese sound pairs that are further apart in terms of perceived phonetic distances, and vice versa. Accordingly, we conducted Category discrimination experiment using AX forced-choice to obtain perceptual discrimination data.

Specifically, this chapter presents two experiments that investigated the ease/difficulty with which Japanese learners perceived two contrasting Chinese sounds. The Category Mapping Study (3.2) empirically investigated the perceived distance between two contrasting Chinese sound categories, and Category Discrimination Study (3.3) investigated the discrimination difficulty of a select pairs to verify the predictions of the model (PAM).

2.2. Category Mapping Study

2.2.1 Methods

2.2.1.1. Participants

Eleven native speakers of Japanese (7 female, 4 male, mean age= 20.8), who had

no experience of learning Chinese, recruited at University of Oregon, participated in the categorical mapping test. Out of all Japanese native speakers, six participants came from areas where standard Tokyo dialect is spoken and the other five were from the western (Kansai) area of Japan. Their average length of stay in the US ranged from 1 to 13 months and the average length is 4.9 months. All reported daily use of Japanese.

2.2.1.2. Materials

Four Mandarin Chinese native speakers (2 male, 2 female, mean age = 26) from mainland China provided speech stimuli for this study. All four speakers were attending the University of Oregon at the time of recording. Their average length of stay in the US was 2.9 years (range: 1.5 years – 4 years). One speaker was from Beijing and the other three were from northern China. The dialects they spoke all belong to the northern dialect family of Chinese, which is very similar to Mandarin Chinese. All reported daily use of Mandarin Chinese.

These four native Mandarin Chinese speakers produced Chinese test consonants and vowels in monosyllabic words placed in a carrier phrase "*qing du___san bian" (Please read ___ three times)*. The test consonants were a full inventory of Mandarin Chinese consonants, including stops [p, p^h, t, t^h, k, k^h], fricatives [s, f, ɕ, x, ʂ], affricates [tɕ, $tɕ^h$, ts, ts^h, tʂ, $tʂ^h$], nasals [ŋ, m, n(onset), n(offset)], ɻ, l (approximants)]. The test vowels were most Mandarin Chinese monophthong vowels and their allophones [a, i, u, ɚ, ə, ɤ, ɑ, y] as well as some diphthongs [ai, ɑu, ou, ɥe, wo, ei]. As for monophthong [u], since it occurred in both open (e.g. [lu] and closed syllable [luŋ], we included stimuli for both phonetic environments.

The consonants were produced in a syllable followed by [u] (i.e., Cu) and the

vowels were produced in a syllable preceded by [l] (e.g., lV). The reason why front vowel [u] was selected as the vowel context for the consonant stimuli was because [u] is more general in terms of vowel environment that co-occur with most consonants in Chinese, while other vowel candidates (e.g. [a] or [i] or [ə]) are very restrictive with regard to the choice of possible proceeding consonants. Similarly, the reason why liquid [l] was selected as the environment for the vowels was also due to the fact that certain vowels (e.g. rounded front vowel [y]) only occur in this consonant environment. Thus choosing [l] as the consonant environment could help to generate a relatively comprehensive list of Chinese vowels produced in the same consonant environment. Having all consonants/vowels produced in the same phonetic environment could further eliminate possible effects introduced by different environment. Since some vowels (e.g. [ə]) only occur in an environment where nasal sound [n] or [ŋ] is the offset in the syllable, these vowels were produced in the following syllable format [l+V+ŋ].

Native Chinese speakers produced all the stimuli syllables embedded in the carrier sentence in two speech rates: a relatively slow rate and a faster rate. The following method was used to elicit slow and fast speech rate. First, the speakers listened to a recording of an isolated target syllable (e.g. [C+u] for the consonants or [l+V+(ŋ)] for the vowels) followed by utterance of a sentence including the word, "*qing du [C+*u*]/ [l+V+(ŋ)] san bian" (Please read [C+u]/ [l+V+(ŋ)] three times)* read in slower speech rate and a faster speech rate. After practicing for a while, the speakers were provided with a list of [C+u] and [l+V+(ŋ)] test syllables written in Chinese orthography with pinyin and tone annotation. All the target syllables were produced in the fourth tone embedded in the carrier sentence twice, and the second production were selected as stimuli.

The speakers were recorded individually in a sound-attenuated booth using a flash digital recorder (Marantz PMD 670) and a standing microphone (SHURE Beta 87) at a sampling rate of 44 kHz and 16-bit quantization. All the syllables containing the target consonant/vowel were then extracted from carrier sentences. This procedure generated a total of 304 stimuli: 38 sounds (23 consonants and 15 vowels examined) x 4 speakers x 2 speech rates). The reading list is attached as Appendix A.

2.2.1.3. Procedure

The 304 speech stimuli (i.e., monosyllabic words containing the test consonants and vowels) were presented to eleven Japanese L2 listeners in two blocks (consonant block and vowel block) using the experiment function of Praat (Paul Boersma, 2002). The order of stimuli presented in each block were randomized. The order of the consonant block and the vowel block was counterbalanced across participants. Japanese listeners were instructed to focus on the consonant portion of the test syllable in the consonant block and the vowel portion of the syllable in the vowel block. They also received a sheet of paper with the trial numbers and a blank space next to each trial number. They were instructed that their first task (identification task) was to write down the sound they just heard using Japanese *hiragana* orthography to identify the closest Japanese sound that sounds like the one they just heard. It is the standard procedure that participants of the L2 sound mapping studies are asked to respond in their L1 (e.g., Guion et al 2000). As *hiragana* script is a syllabary, the participants in effect wrote a syllable that they heard. Immediately after the identification response, they were also asked to rate the goodness-of-fit of the Chinese sound to the selected Japanese consonant/vowel category on a scale ranging from 1 to 7. The score of 1 indicated poor exemplar while 7

indicated very good exemplar.

2.2.2. Results

The results are presented in Table 2.1 and 2.2. Table 2.1 presents how Chinese obstruents [p, p^h, t, t^h, k, k^h, tɕ, $tɕ^h$, ts, ts^h, s, f, ɕ, tʂ, $tʂ^h$, ʂ] were mapped to Japanese sounds, while Table 2.2 presents the how Chinese sonorants [ŋ, ɻ, m, n(onset), n(offset), l] were mapped to Japanese sounds. The Chinese test sounds are represented vertically with the label on the first column, and the selected Japanese sounds are represented horizontally. There are two rows of response data for each Chinese sound category. The numbers on the first row indicate the average goodness rating. The numbers on the second row indicate the % frequency with which this particular Japanese category was selected as the closest counterpart for the Chinese consonant stimuli they have heard. Numbers in bold indicate the modal identification, in other words, the most frequently selected Japanese category.

Table 2.1. Goodness rating and mean percent identification of Chinese obstruent stimuli in terms of Japanese categories. Boldfaced values indicate the modal identification response. The goodness ratings are based on a scale that ranged from "poor exemplar"(1) to "very good exemplar" (7).

Percent Identification and Rating															
Consonant (Obstruents)	**ぶ b**	**ぷ p**	**ど d**	**と t**	**ぐ g**	**こ k**	**ふ ɸ**	**す s**	**しゅ ɕ**	**つ ts**	**ず dz**	**じゅ dʑ**	**ちゅ tɕ**	**る ɾ**	**ゆ j**
p	4.1	**4.4**				3.0	2.7								
	24	**72**				1	3								
p^h		**4.8**		2.3			3.0		4.0	2.7					
		85		3			2		1	8					
t			3.3	**3.6**						2.7	5		4.0	3.0	
			16	**77**						3	1		1	1	
t^h				**3.1**						2.9					
				91						9					

Table 2.1. (continued).

Percent Identification and Rating																
Consonant (Obstruents)	**ぶ b**	**ぷ p**	**ど d**	**と t**	**ぐ g**	**こ k**	**ふ ɸ**	**す s**	**しゅ ɕ**	**つ ts**	**ず dz**	**じゅ dʑ**	**ちゅ tɕ**	**る ɾ**	**ゆ j**
k		3			3.7	**4.2**			3						
		1			25	**73**			1						
k^h					3.5	**4.2**									
					2	**98**									
f		2.8		2.3			**2.6**	2	1	2			1.7		
		22		13			**49**	2	2	9			3		
x		1.0			2.0	3.3	**3.2**							5.0	
		1			1	34	**60**							1	
s		3.0				1.0	2.5	**3.4**	1.0	3.7					
		5				1	2	**83**	2	7					
ʂ		2.6		1.9				2.9	**2.7**	1.6			2.9		
		10		11				14	**41**	8			16		
ɕ		3.3	1	3					**4.6**	1.5			3.4		1.0
		3	1	1					**75**	2			16		1
ts	5.0	3.0	1.0	2.0			1.5	1.0		**4.5**	3.6		4.0		
	1	1	1	1			2	1		**77**	14		1		
ts^h				2.8				2.5		**3.8**	2		1.0		
				9				9		**80**	1		1		
tʂ			1.0	2.7			2.0			2.6		1.8	**2.8**	1.0	
			1	23			1			33		5	**36**	1	
$tʂ^h$				2.6		3.0			1.0	2.3			**2.8**		
				28		1			1	30			**40**		
tɕ				1.5					5.0	3.0		3.5	**4.0**		
				2					1	3		17	**76**		
$tɕ^h$										3.0		4.0	**4.2**		
										1		1	**98**		

We can see in Table 2.1, Chinese aspirated stops [p^h, t^h, k^h] were identified mostly as Japanese voiceless stops [p, t, k], and the identification rates (85, 91, and 98) and

goodness ratings (4.8, 3.1, and 4.2) were high. However, Chinese unaspirated stops [p, t, k] were also categorized primarily into Japanese voiceless stops with lower but fairly high identification rates (72, 77, and 73) and goodness ratings (4.4, 3.6, and 4.2). It is noted that unaspirated stops [p, t, k] were categorized into a larger range of Japanese categories compared to aspirated stops. For instance, unaspirated stop [t] was identified as six Japanese consonant categories [d, t, ts, tɕ, dz, r], while its aspirated counterpart [t^h] was identified as two Japanese consonants [t] and [ts] with [ts] as the modal (most frequent) classification for both sounds. Similarly, unaspirated stop [k] was identified as Japanese category [p, g, k, ɕ], while aspirated stop [k^h] was only identified as two Japanese categories [g] and [k] with [k] as the modal classification for both sounds. Japanese listeners seemed to have encountered more difficulties in reaching a consensus regarding the closest Japanese category for Chinese [p]-[p^h] pair, since both sounds were classified into a large number of Japanese categories. Aspirated stop [p^h] was identified into Japanese stop, fricative and affricate categories---[p, t, ɸ, ɕ, ts], while unaspirated [p] was also identified into several Japanese stops and fricative category---[b, p, k, ɸ]. These results suggested that Chinese aspirated and unaspirated stops may be difficult for Japanese learners to discriminate. In particular, Japanese listeners seemed to be having a challenge characterizing Chinese unaspirated stops with existing Japanese consonant categories.

Chinese fricatives [f, x, s, ʂ, ɕ] were identified mostly as Japanese fricatives [ɸ, s, ɕ]. However, the identification rates and goodness rating varied greatly for each sound. [f, x] were both identified as Japanese bilabial fricative [ɸ] most frequently, while the identification rates were not high (49 and 60) and the goodness ratings were moderate

(2.6) and relatively high (3.2) respectively. Besides this modal identification category [ɸ], [f] was also identified as a wide range of Japanese consonants [p, t, s, ɕ, ts, and tɕ], while [x] was also identified as Japanese consonants [p, g, k, r]. These results showed that although Chinese [x] and [f] were categorized into the same Japanese category most frequently, Japanese leaners did seem to sense differences between these two sounds, since they were secondarily identified into two different sets of Japanese consonant categories. Japanese L2 learners did not seem to have encountered great difficulty in discriminating these two sounds.

Chinese fricative [s] was identified mostly as Japanese fricative [s] with high identification rates (83) and high goodness rating (3.4). Fricatives [ʂ, ɕ] were both identified as Japanese fricative [ɕ]. However, the identification rate and the goodness rating for [ʂ] were 41 and 2.7, while those for [ɕ] were higher, at 75 and 4.6. Japanese does not have retroflex sound [ʂ], and the results here showed that Japanese listeners did have difficulty finding a good counterpart in Japanese sounds. These mapping results for Chinese fricatives indicated that the discrimination between some of the fricatives may be easy, e.g. between [s] vs. [ɕ] which were classified into different Japanese categories, or modestly easy, e.g., between [ʂ] vs. [ɕ] which were classified into a single Japanese category but with different goodness ratings varied. However, the discrimination could be difficult for some other fricative pairs, especially for those pairs which were identified into the same Japanese categories with similar identification rates and goodness rating (e.g. [f] and [x]).

As for affricates, Chinese aspirated [ts^h, tʂ^h, tɕ^h] and unaspirated affricates [ts, tʂ, tɕ] were mostly identified as Japanese voiceless affricates [ts, tɕ]. More specifically, [ts^h,

ts] were both classified as the same Japanese voiceless affricate category [ts] with high identification rates (77, 80) and high goodness ratings (4.5, 3.8). Both [ts^{h}, ts] were also categorized into a range of Japanese consonant categories [t, d, s, b, p], including Japanese voiced affricate [dz]. The sounds [ts^{h}, ts] were the only affricate pair that was classified into Japanese voiced affricate [dz], although the identification rates were fairly low (1, 14).

Two Chinese affricate pairs [tʂ, tʂ^{h}] and [tɕ, tɕ^{h}] were categorized differently from [ts^{h}, ts]. Unlike [ts^{h}, ts], which were mapped to Japanese [ts], Chinese [tʂ, tʂ^{h}] and [tɕ, tɕ^{h}] were mapped to the same Japanese voiceless affricate [tɕ], with different frequencies and goodness-of-fit rating. The retroflex [tʂ, tʂ^{h}] pair was identified with low identification rates (36, 40) and low goodness rating (2.8, 2.8), while the alveo-palatal [tɕ, tɕ^{h}] pair was identified with fairly high identification rates (76, 98) and high goodness rating (4.0, 4.2), consistent with the pattern we found for fricatives. Similar to the tendency observed for stops, Chinese unaspirated affricates were identified into a larger range of Japanese categories than their aspirated affricate counterparts, which implied that Japanese native listeners had more difficulty in consistently categorizing Chinese unaspirated affricates than aspirated ones.

As indicated in Table 2.2, all six Chinese sonorants [l, ɻ, m, n (onset), n (offset), ŋ] were consistently (frequency > 75%) classified as the exemplar of one particular Japanese category, i.e., [ɾ] for [l, ɻ], [n] for onset [n[, and [N] for offset [n], and [N] for [ŋ]. The goodness ratings received by these six sonorants for their corresponding classified Japanese category were also fairly high (4.7, 4.4, 4.2, 4.3, 5.0, 5.0). On one hand, Chinese nasals [m] and [n (onset)] were classified separately as Japanese nasal [m]

and [n] almost all the time (the identification rates were 1 and 0.99 respectively). On the other hand, a problematic case was found with [n (offset)] and [ŋ], both of which were classified into the same Japanese category [N] with high identification rates [78, 92] and high goodness rating (5.0, 5.2). The pair [l, ɻ,] was also classified into a single Japanese [r] with similar identification rates (95, 80) and goodness ratings (4.7, 4.4). These results indicated that while discriminating Chinese nasals [m, n (onset)] is likely to be easy for Japanese native speakers, discriminating between Chinese [l, ɻ,] sonorant pair and [n(offset), ŋ] pair might not be as easy since these two pairs were both classified into the same Japanese category with similar frequency and goodness-of-fit ratings.

Table 2.2. Goodness rating and mean percent identification of Chinese sonorant stimuli in terms of Japanese categories. Boldfaced values indicate the modal identification response. The goodness ratings are based on a scale that ranged from "poor exemplar"(1) to "very good exemplar" (7)

Consonants (Sonorants)	る **ɾ**	む **m**	ぬ **n**	ん **N**	ぶ **b**	ず **dz**	ぐ **g**	ふ **ɸ**	じゅ **dʑ**	**x None**
l	**4.7**	3.0	3.3							
	95	1	3							
ɻ,	**4.4**	3.0			2.0	2.5	3.0	5.0	2.0	
	80	2			1	2	1	1	13	
m		**4.2**								
		100								
n(onset)		2.0	**4.3**							
		1	**99**							
n(offset)				**5.0**						3.8
				78						22
ŋ				**5.2**						3.0
				92						8

The results for the category mapping between Chinese and Japanese vowels are presented in Table 2.3, 2.4 and 2.5. Table 2.3 presents the results for the front vowels [i, y, ei, ɥe], Table 2.4 presents the results for central vowels [a, ɚ, ə, ai] and Table 2.5 presents the results for back vowels [ɑ, ɤ, u, u in closed syllable, ɑu , uo, wo].

Table 2.3. Goodness rating and mean percent identification of Chinese front vowel stimuli in terms of Japanese categories. Boldfaced values indicate the modal identification response. The goodness ratings are based on a scale that ranged from "poor exemplar"(1) to "very good exemplar (7).

Vowels (front vowels)	**い i**	**う ɯ**	**あ a**	**え e**	**えい ei**	**えあ ea**	**えう eɯ**	**いあ ia**	**いえ ie**	**いう iɯ**	**お o**	**うあ ɯa**	**うえ ɯe**	**うい ɯi**	**うー ɯ ɯ**
i	**4.8**							2.0	3.0	2.0	7.0				
	90							2	1	1	1				
y	3.3	**3.3**							3.7	2.5					3.0
	38	**60**							3	2					1
ei	3.9			3.1	**4.4**		1.0							2.0	
	17			11	**70**		1							2	
ɥe	2.0	**2.8**	1.6	3.2		2.0		2.0	2.6			2.1	2.8		
	2	**30**	8	1		2		1	18			17	14		

Table 2.4. Goodness rating and mean percent identification of Chinese central vowel stimuli in terms of Japanese categories. Boldfaced values indicate the modal identification response. The goodness ratings are based on a scale that ranged from "poor exemplar"(1) to "very good exemplar" (7).

Vowels	**あ a**	**あえ ae**	**あい ai**	**あう aɯ**	**え e**	**えい ei**	**い i**	**お o**	**う ɯ**	**うあ ɯa**	**うあい ɯai**	**うえ ɯe**	**うえい ɯei**
ai	1.6	2.8	**3.9**	2.0	3.0	4.0	3.4	5.0	2.0	2.0	2.5	1.0	5.0
	6	6	**35**	1	32	2	8	1	2	1	2	1	2

Table 2.4. (continued)

vowels (central vowel)	あ a	あー aa	あーる aarɯ	あーと aato	お o	あん aN	あお ao	ある arɯ	あう aɯ	え e	えう eɯ	い i	ん N	おん oN	おう oɯ	う ɯ	うあ ɯa	うあー ɯaa
a	**4.0**						1.0		2.0	2.6		2.0					2.2	2.0
	84						1		1	9		1					3	1
ɐ	**3.6**	3.6	3.3	5.0		4.0		2.8	2.2									
	57	25	3	1		2		6	6									
ə	2.5				**3.0**					2.6	2.0	1.0	2.0	5.0	2.8	2.3	3.0	
	11				**30**					27	1	1	1	1	2	26	1	

Table 2.5. Goodness rating and mean percent identification of Chinese back vowel stimuli in terms of Japanese categories. Boldfaced values indicate the modal identification response. The goodness ratings are based on a scale that ranged from "poor exemplar"(1) to "very good exemplar" (7).

vowels(back vowels)	あ a	う ɯ	お o	あお ao	おう oɯ	うあ ɯa	うお ɯo	うえ ɯe	うおう ɯoɯ	うおうあ ɯoɯa	うーあ ɯɯa	あう aɯ	えい ei	おあ oa	おえ oe	おん on
	3.3		2.5	3.4	2.8	1.5						3.7			3.0	
ɑ	**60**		19	8	6	2						3			1	
	2.5	**3.2**	1.0			2.7	1.5	2.7			2.0			3.0		
ɤ	9	**41**	1			38	2	7			1			1		
	3.0	**3.8**	1.7		3.0											
u	1	**95**	3		1											
u in closed syllable	2.0	1.5	**3.3**		3.3											5.0
	1	1	**94**		3											1
	2.4	3.0	3.3	**4.2**	2.9				2.0			2.8	4.0			
ɑu	8	2	33	**34**	16				1			5	1			
	5.0	3.1	4.0		**3.7**		3.0					3.0		3.0		
ou	1	11	35		**48**		2					1		1		
	2.5	3.0	3.0	4.0		**3.1**	3.2	2.5		4.0				3.6	3.0	
wo	9	18	5	1		**48**	7	5		1				6	1	

As indicated in Table 2.3, unrounded front vowel [i] and rounded front vowel [y] were classified into two different Japanese vowel categories [i] and [ɯ] respectively. Unrounded [i] seemed to be a good exemplar for Japanese [i] since the identification rate was (90) with goodness rating as high as (4.8). Out of these four Chinese front vowels, only [i] was consistently (frequency > 75%) classified as the exemplar of one particular Japanese category [i]. The identification rate for rounded [y], however, was much lower (60) with moderate goodness rating (3.3). Diphthongs [ei] and [ɥe] were classified into one or two Japanese vowel categories. For diphthong [ɥe], the Japanese category it was most frequently classified into was a Japanese single vowel [ɯ]; however, the identification rate was only 30 percent and goodness rating was not high (2.8). In contrast, diphthong [ei] was most frequently classified into a Japanese two-vowel combination [ei] with high identification rate (70) and goodness-of-fit rating (4.4). These results suggested that Japanese native speakers did not have reliable perceptual mapping for Chinese [y] and [ɥe] using Japanese categories.

Table 2.4 presents the results for Chinese central vowels, including three monophthongs [a, ɐ, ə] and one diphthong [ai]. Both [a ɐ] were most frequently classified into the single Japanese vowel [a]. However, Chinese [a] seemed to be a better exemplar of Japanese vowel [a], because of its higher identification rate [84] and goodness rating [4.0] compared to 57 and 3.6 for [ɐ]. Chinese [ə] was classified into a wide range of Japanese vowels [a, o, e, i, eɯ, N, oN, oɯ, ɯ, ɯa] and the most frequency identified category [o] only had identification rate as low as 30. Diphthong [ai] was also classified into a number of different Japanese vowel categories. The most frequently identified Japanese category was two-vowel combination [ai] with a low identification

rate of 35, which was only slightly higher than that of the second most frequently identified Japanese single vowel category [e] (32). These results indicated that Japanese speakers did not agree on how Chinese central vowel [ə] and diphthong [ai] mapped to Japanese vowel categories.

The results for six Chinese back vowels (four monophthongs [ɑ, ɤ, u in open syllable, u in closed syllable] and three diphthongs [ɑu, ou, wo]) were presented in Table 2.5. Out of these six back vowels, only [u] occurring in both open and closed syllable environments was consistently (frequency > 75%) classified as an exemplar of a particular Japanese category. However, [u] in open and closed syllable were identified as two different Japanese vowel categories. Chinese [u] occurring in open syllable (e.g. [nu]) was classified as Japanese vowel [ɯ] 95% of the time, while Chinese [u] occurring in closed syllable (e.g. [luŋ]) was classified as Japanese vowel [o] 94% of the time. In addition to Chinese [u] occurring in open syllable, [ɤ] was also classified into Japanese category [ɯ] most frequently, yet with a much lower classification rate 41. Similarly, Japanese speakers did not seem to have a consensus on the closest Japanese categories three diphthongs [ɑu, ou, wo] should be classified into, the most frequently identified Japanese categories were respectively [oɯ, ɯa, ɯo] and none of their identification rates (34, 48, 48) exceeded 50. These results suggest vowels that Japanese speakers had great difficulty in identifying a single Japanese vowel category for Chinese vowels there were not present in Japanese vowel inventory (e.g. mid-back vowel [ɤ] and diphthongs [ɑu, ou, wo]).

In order to assess the perceived phonetic distance between Chinese category and classified Japanese category, it is necessary to take into consideration of both rate with

which a Japanese category is selected as a close counterpart to a Chinese sound category, and the degree of goodness-of-fit of the Chinese sound to the Japanese counterpart. We used "fit index" proposed by Guion et al (2000) as the measuring metric for this purpose. This metric is calculated by multiplying the percentage of identification and the value of goodness rating. For instance, the fit index of Chinese unaspirated stop [p] mapped to Japanese voiceless stop category [p] was 3.17, which was obtained by multiplying the proportion of [p] identification (72) with its corresponding mean goodness rating (4.40). Thus for each Chinese sound, the higher fit index it receives, the better exemplar it is of the identified Japanese category.

The fitness indexes derived for Chinese consonants and vowels examined above are reported in Table 2.6 and Table 2.7. There is a large range of fit index, spanning from 0.86 to 4.74 for consonants, and 0.75 to 4.74 for vowels. Based on these fit values, we separated Chinese sounds into three groups, good, fair and poor exemplar so that each group evenly represent one third of the range. For instance, Chinese consonants with fit index falling into the bottom one third of the entire range (from 0.86 to 2.15) were labeled as poor exemplars, middle one third of the entire range (from 2.16 to 3.44) were labeled as fair exemplars, and top one third of the entire range (from 3.45 to 4.74) were labeled as good exemplars of the corresponding Japanese category. We recognize that this grouping is not done using qualitative benchmarks. However, given that there is no established criterion or qualitative benchmarks for fit index, we decided to group the sound categories based on the data quantitatively.

Fit index measures above allowed us to access the phonetic distances between each Chinese sound and the Japanese sound it was categorized into. However, in order to

make predictions on discrimination difficulties of two Chinese sounds under PAM framework, we need to compare this Chinese sound pair in terms of six assimilation patterns discussed in Introduction ((1) Two-Category Assimilation (CC Type); (2) Category-Goodness Difference (CG Type); (3) Single-Category Assimilation (SC Type); (4) Both Uncategorizable (UU Type); (5) Uncategorized versus Categorized (UC Type); (6) Nonassimilable.) Predicted discrimination difficulty of two sounds based on PAM framework were reported in the rightmost column in Table 2.6 and 2.7. For consonant pairs, for instance, Chinese aspirated stop [t^h] and unaspirated stop [t] were both categorized into the same Japanese voiceless stop category [t] as "fair" exemplars (see Table 2.6), this case falls into the SC Type in which these two sounds were perceived to be close to each other in terms of phonetic distance. Accordingly, the discrimination difficulty between Chinese [t^h] and [t] for Japanese L2 learners is predicted to be hard. Similarly, the difficulty of discriminating between Chinese affricates [$tɕ^h$] and [tɕ] were predicted to be "moderate" because these two sounds were categorized into the same Japanese category of [tɕ] respectively yet the former was labeled as "good" and the latter was labeled as "fair" exemplar (CG Type).

Fit indexes and predicted discrimination difficulty of Chinese vowel pairs were reported in Table 2.7. For instance, Chinese unrounded front vowel [i] was categorized into Japanese category [i] as good exemplar, while Chinese rounded vowel [y] was not categorizable into any Japanese vowel category since it was perceived to fall in between two Japanese categories [ɯ] and [i]. For a specific Chinese sound, we followed the standard used in previous research (Guion et al, 2000) to include Japanese categories that have identifications that were more than 30% as identified categories. If one Chinese

Table 2.6. Fit indexes derived for Chinese consonants in terms of Japanese categories. The fit index was computed as the multiplication of proportion of identifications and goodness ratings. Only identifications that were more than 30% are included. "Good", "Fair" and "Poor" exemplars are labeled based on fit index and respectively represent the bottom, middle and top 1/3 of the entire range. The predicted discrimination difficulty of two sound pairs ("Easy", "Moderate", "Hard") was reported in the rightmost column.

Chinese Consonants	Most Frequently Identification	Portion of identifications	Goodness Rating	Fit Index		Discriminability Prediction
[pʰ]	[p]	0.85	4.75	4.04	good [p]	Moderate
[p]	[p]	0.72	4.40	3.17	fair [p]	
[tʰ]	[t]	0.91	3.13	2.85	fair [t]	Hard
[t]	[t]	0.77	3.60	2.77	fair [t]	
[kʰ]	[k]	0.98	4.22	4.14	good [k]	Moderate
[k]	[k]	0.73	4.22	3.08	fair [k]	
[tsʰ]	[ts]	0.80	3.76	3.01	fair [ts]	Moderate
[ts]	[ts]	0.77	4.49	3.46	good [ts]	
[tɕʰ]	[tɕ]	0.98	4.17	4.09	good [tɕ]	Moderate
[tɕ]	[tɕ]	0.76	4.04	3.07	fair [tɕ]	
[tʂʰ]	[tɕ]	0.40	2.83	1.13	poor [tɕ]	Hard
	[ts]	0.30	2.27	0.68	poor [ts]	
[tʂ]	[tɕ]	0.36	2.75	0.99	poor [tɕ]	
	[ts]	0.33	2.62	0.86	poor [ts]	
[s]	[s]	0.83	3.42	2.84	fair [s]	Easy
[ʂ]	[ɕ]	0.41	2.67	1.09	poor [ɕ]	
[ɕ]	[ɕ]	0.75	4.60	3.45	good [ɕ]	Easy
[ʂ]	[ɕ]	0.41	2.67	1.09	poor [ɕ]	
[x]	[h]	0.63	3.15	1.98	poor [h]	Easy
	[k]	0.34	3.33	1.13	poor [k]	
[f]	[h]	0.49	2.63	1.29	poor [h]	
[m]	[m]	1.00	4.22	4.22	good [m]	Easy
[n onset]	[n]	0.99	4.32	4.28	good [n]	
[ŋ]	[N]	0.92	5.15	4.74	good [N]	
[n offset]	[n]	0.78	5.02	3.92	good [n]	
[l]	[ɾ]	0.95	4.74	4.50	good [ɾ]	Hard
[ɻ]	[ɾ]	0.80	4.41	3.53	good [ɾ]	

Table 2.7. Fit indexes derived for Chinese vowels in terms of Japanese categories. The fit index was computed as the multiplication of proportion of identifications and goodness ratings. Only identifications that were more than 30% are included. "Good", "Fair" and "Poor" exemplars are labeled based on fit index and respectively represent the bottom, middle and top 1/3 of the entire range. The predicted discrimination difficulty of two sound pairs ("Easy", "Moderate", "Hard") was reported in the rightmost column.

Chinese Vowels	Most Frequently Identification	Portion of identifications	Goodness Rating	Fit Index		Discriminability prediction
[i]	[i]	0.94	4.77	4.48	good [i]	Easy
[y]	[ɯ]	0.56	3.29	1.84	poor [ɯ]	
	[i]	0.38	3.33	1.27	poor [i]	
[y]	[ɯ]	0.56	3.29	1.84	poor [ɯ]	Easy
	[i]	0.38	3.33	1.27	poor [i]	
[u]	[ɯ]	0.95	3.79	3.6	good [ɯ]	
[u]	[ɯ]	0.95	3.79	3.6	good [ɯ]	Easy
[ɤ]	[ɯ]	0.41	3.17	1.3	poor [ɯ]	
	[ɯa]	0.38	2.67	1.01	poor [ɯa]	
[a]	[a]	0.84	3.97	3.33	good [a]	Moderate
[ɚ]	[a]	0.57	3.62	2.06	fair [a]	
[ɑ]	[a]	0.61	3.63	2.23	fair [a]	Easy
	[o]	0.2	2.6	0.52	poor [o]	
[ə]	[o]	0.3	3.04	0.91	poor [o]	
[ə]	[o]	0.3	3.04	0.91	poor [o]	Moderate
[u in diphthong]	[o]	0.94	3.26	3.06	fair [o]	
[u in diphthong]	[o]	0.94	3.26	3.06	fair [o]	Easy
[ou]	[oɯ]	0.48	3.76	1.8	poor [oɯ]	
	[o]	0.35	4	1.4	poor [o]	
[wo]	[ɯa]	0.48	3.12	1.5	poor [ɯa]	Easy
[ɑu]	[ao]	0.34	4.23	1.44	poor [ao]	
	[o]	0.33	3.34	1.1	poor [o]	
[ai]	[ai]	0.35	3.94	1.38	poor [ai]	Easy
	[e]	0.32	2.96	0.95	poor [e]	
[ɥe]	[ɯ]	0.27	2.79	0.75	poor [ɯ]	
[ei]	[ei]	0.68	4.37	2.97	fair [ei]	Easy
[ai]	[ai]	0.35	3.94	1.38	poor [ai]	
	[e]	0.32	2.96	0.95	poor [e]	

sound had more than one identified Japanese categories, this sound does not assimilate to any single Japanese category and thus was redeemed as "uncategorizable.". The assimilation pattern of Chinese [i] and [y] matches PAM's UC Type, thus the difficulty of discriminating these two sounds was predicted to be "easy". In the same vein, the discrimination between sound pairs [u] vs. [y], [ɑ] vs. [ə], [u] vs. [ɤ] all fall into UC Type and were thus predicted to be "easy" as well. The difficulty in discriminating [a] vs. [ɚ] pair, however, was predicted to be moderate since Chinese [a] and [ɚ] were both categorized into Japanese vowel category [a] as "good" and "fair" exemplar respectively, which falls into CG Type with moderate perceived phonetic distances.

As shown above, the results of mapping study enabled us to make predictions regarding the difficulty discriminating these L2 Chinese sounds for Japanese learners. These predictions were tested with a discrimination experiment in the next section.

2.3. Category Discrimination Study

The purpose of this experiment was to examine to what degree Japanese L2 learners could discriminate Chinese consonant and vowel pairs with varying perceived phonetic distance. The mapping study helped us to categorize Chinese consonants and vowel sounds as "good", "moderate" and "poor" exemplars for classified Japanese categories using fit index measures. These results enabled us to make predictions on discrimination difficulties of the Chinese sounds for Japanese L2 learners. In this section, discrimination of select Chinese sounds by Japanese L2 learners were empirically examined to test the prediction derived from PAM.

2.3.1 Methods

2.3.1. 1. Participants

Ten native speakers of Japanese (7 female, 3 male, mean age= 21.9) (See Table 2.8) who had varying experience of learning Chinese, recruited at University of Oregon, participated in this discrimination experiment. None of these participants participated in the Category Mapping Study. Out of all Japanese native speakers, three participants came from areas where standard Tokyo dialect is spoken and the other seven were from Kansai area of Japan. Their average length of stay in the US ranged from 1 to 60 months and the average length is 14 months (See Table 2.8). The ages when these participants started to learn Chinese (AOL) ranged from 6 to 24 years old, with the mean 17.5 years old. All participants except one started learning Chinese before age 15. One participant started to learn Chinese at the age of 6. The length of Chinese learning experience ranged from 3 months to 60 months, with the mean duration of learning Chinese for 28.5 months. Most of them had studied Chinese as a second language in college. None of ten participants had been to China and all participants reported daily use of Japanese.

Table 2.8. Language background of Japanese L2 Participants

	Japanese L2 Learners of Chinese (n = 10)
Age (years)	21.9 (19-27)
AOL (years)	17.5 (6-24)
Chinese instruction (months)	28.5 (3-120)
Length of stay in US (months)	14 (1-60)

2.3.1.2. Materials

Ten consonant pairs (Table 2.9) and five vowel pairs (Table 2.10) were selected for the discrimination test to represent all levels of predicted difficulty (7 "easy", 5 "moderate" and 3 "hard"). The predicted difficulty levels presented in the table were derived in the analysis of the previous section.

Table 2.9. Consonant pairs selected for discrimination test

Consonant Pairs	Chinese Contrast Pairs	Classified Japanese Category	Predicted Difficulty
1	[k^h] vs. [k]	good [k], fair [k]	Moderate
2	[p^h] vs. [p]	good [p], fair [p]	Moderate
3	[t^h] vs. [t]	fair [t], fair [t]	Hard
4	[ts^h] vs. [ts]	good [ts], fair [ts],	Moderate
5	[$tʂ^h$] vs. [ts]	Uncategorizable, good [ts]	Easy
6	[$tʂ^h$] vs. [tʂ]	Uncategorizable, Uncategorizable,	Hard
7	[$tɕ^h$] vs. [tɕ]	good [tɕ], fair [tɕ]	Moderate
8	[ɕ] vs. [ʂ]	good [ɕ], poor [ɕ]	Easy
9	[x] vs. [f]	uncategorizable, poor [h]	Easy
10	[l] vs. [ɻ,]	good[ɾ], good [ɾ]	Hard

Table 2.10. Vowel pairs selected for discrimination test

Vowel Pairs	Chinese Contrast Pairs	Classified Japanese Category	Predicted difficulty
11	[i] vs. [y]	good[i] vs. uncategorizable	Easy
12	[a] vs. [ɚ]	good [a], fair [a]	Moderate
13	[u] vs. [y]	poor [u], uncategorizable	Easy
14	[ɑ] vs. [ə]	Uncategorizable, poor [o]	Easy
15	[u] vs. [ɤ]	good[ɯ], uncategorizable	Easy

2.3.1.3. Procedures

An AX forced-choice discrimination test was used for the category discrimination test. In this test, listeners heard one pair of sounds per trial and were asked to indicate whether they were the same (press "1") or different (press "2"). For each contrasting sound pairs AX, there were 4 pairs of “same (or catch)" trials (2 repetitions of AA pairs,

2 repetitions of XX pairs), and 4 pairs of "different" trials with switched sound positions (2 repetition of AX pairs and 2 repetition of XA pairs). For consonant pair [k^h] and [k], for example, there were 2 repetitions of [k^h] vs. [k^h] trials, 2 repetitions of [k] vs. [k] trials, 2 repetitions of [k^h] vs. [k] trials, and 2 repetitions of [k] vs. [k^h] trials. Accordingly, there were a total of 480 trials (including 320 consonant and 160 vowel pairs): 15 pairs (10 consonant pairs and 5 vowel pairs) × 2 speech rates × 2 speaker sexes × 8 types (2 AA pairs, 2 XX pairs, 2AX pairs, 2 XA pairs). This experiment was divided into two blocks---consonant block and vowel block. In each block, before entering the real test session, there was a practice test in which the participants practiced with two pairs of sounds to familiarize themselves with the experiment procedure. The order of consonant and vowel blocks was counterbalanced across the participants.

The participants were tested individually in a sound booth and heard these sound pairs in randomized order within each block. In each block, Japanese listeners were told to ignore individual differences of these sound pairs and indicate whether they are different or the same by pressing "1" or "2". This categorical discrimination test lasted approximately 40 minutes.

2.3.1.4. Analysis

In order to provide an unbiased measure of listeners' responses to sound pair contrast, we need to take into consideration of responses to both different and catch trials. The metric we used here is called A-prime (A') score, A-prime scores for each of the Chinese sound pair contrast examined was by multiplying the proportion of "hits" (correct responses in "different" trials) and "false alarms" (incorrect responses in "catch" trials) calculated (Snndgrass, Levy—Berger, & Haydon, 1985).

2.3.1.5. Results

A-prime scores are reported in Table 2.11 below. Higher values of A-prime score indicate higher sensitivity or discrimination performance. The A-prime scores for these fifteen contrast pairs ranged from 0.65 to 0.91, and contrast pairs are listed in Table 2.11 based on their A prime scores in descending order. The predicted discriminability for each sound contrasts by PAM was then evaluated in comparison to its A prime score ranking and labeled as either "different", "consistent" or "almost consistent". For instance, the discrimination difficulty of contrast pair [a] vs. [ɚ] was predicted to be "Moderate" by PAM, which was supposed to be associated with an A prime score that ranked at the middle among the group of all A prime scores. However, it turned out to have the highest A prime score (0.91) and thus was labeled as "different". Contrasting pair [t^h] vs. [t] which was predicted to be "difficult" to discriminate by PAM, was found to have an A prime score of 0.73 which was positioned at the bottom of the A prime score group, thus was deemed as "consistent" with the prediction. There are also two contrast pairs ([ɑ] vs. [ə] and [$tʂ^h$] vs. [tʂ]) which were label as "almost consistent". Contrast [ɑ] vs. [ə] was predicted to be "easy" to discriminate and its prime score (0.80) ranked in between high and middle of the group, thus it was labeled as "almost consistent". Similarly, for contrast [$tʂ^h$] vs. [tʂ], the discriminablity of this pair was predicted to be "hard", and its A prime score was positioned at the dividing line between bottom and middle range of the group. Thus it was deemed as "almost consistent" with the prediction. This results of this evaluation are reported in the last column in Table 2.11.

Table 2.11. Comparison between predicted difficulty and difficulty based on A prime scores of 15 contrast pairs. Discriminability that were inconsistent with the prediction were labeled as "different" in bold.

Sound Pair	**PAM categories**	**PAM prediction**	**A' score**	**Comparison**
[a] vs. [ɔ]	good [a], fair [a]	Moderate	0.91	**different**
[i] vs. [y]	good[i] vs. uncategorizable	Easy	0.90	consistent
[tʂʰ] vs. [ts]	Uncategorizable, good [ts]	Easy	0.90	consistent
[tsʰ] vs. [ts]	good [ts], fair [ts],	Moderate	0.89	**different**
[u] vs. [y]	poor [u], uncategorizable	Easy	0.88	consistent
[u] vs. [ɤ]	good[ɯ], uncategorizable	Easy	0.86	consistent
[ɑ] vs. [ə]	Uncategorizable, poor [o]	Easy	0.80	almost consistent
[pʰ] vs. [p]	good [p], fair [p]	Moderate	0.80	consistent
[tɕʰ] vs. [tɕ]	good [tɕ], fair [tɕ]	Moderate	0.79	consistent
[ɕ] vs. [ʂ]	good [ɕ], poor [ɕ]	Easy	0.78	**different**
[kʰ] vs.[k]	good [k], fair [k]	Moderate	0.77	consistent
[tʂʰ] vs. [tʂ]	Uncategorizable, Uncategorizable,	Hard	0.76	almost consistent
[x] vs. [f]	uncategorizable, poor [h]	Easy	0.75	**different**
[tʰ] vs. [t]	fair [t], fair [t]	Hard	0.73	consistent
[l]vs. [ɻ]	good[ɾ], good [ɾ]	Hard	0.65	consistent

Based on the criteria above, the evaluation revealed that the "actual" discrimination difficulties for nine contrast pairs out of fifteen, determined based on A-prime score rankings were found to be consistent with the predictions in the category

mapping study. Two contrast pairs were deemed as "almost consistent" with the prediction since their A-prime scores fell at the dividing lines of two sections. Nevertheless, there are four contrast pairs (highlighted in bold in Table 2.11) for which the results turned out to be different from the prediction. This discrepancy was found for four consonant contrast pairs and one vowel pair, and showed two different patterns.

One type of discrepancy occurred when PAM predicts the contrast pair to be "moderate" to discriminate, the high A prime score indicated that it was relatively "easy" instead. The vowel pair [a] and [ɚ] and affricate pair [ts^h] and [ts] both fell into this type. Sounds in both contrast pairs were respectively determined as "good" ([ts^h] and [a]) and "fair" ([ts] and [ɚ]) exemplars for Japanese consonant [ts] and vowel [a] in the category mapping study. This makes CG type (PAM) and it predicts "moderate" discriminability. However, the results of this discrimination experiment indicated that it was quite easy to distinguish this pair, since A prime scores for these two pairs were as high as 0.91 and 0.89. Another type of discrepancy occurred when the category mapping results showed the contrasting sound pair was not easy to discriminate although the predicted discrimination difficulty was "easy". Two consonant pair [ɕ] vs. [ʂ], and [x] vs. [f] belonged to this type. Fricative pair [ɕ] and [ʂ] were respectively "good" and "poor" exemplars for Japanese category [ɕ], and the discrimination of these sounds were predicted to be easy. However, the results here indicated the discrimination showed moderate difficulty. Fricative [f] was both perceived to be "poor" exemplar for Japanese consonant category [h] while the contrasting sound [x] was not categorizable because it was heard as falling in between two Japanese categories (Japanese [h] and [k] categories for Chinese fricative [x]). PAM predicts the discrimination between sound pairs of UC

type are "easy". However, the results of this experiment showed these two sound pairs were actually harder to distinguish. It is interesting to note that both types of discrepancies discussed above involved adjacent discrimination difficulty groups (e.g. "Moderate" to "Easy", and "Easy" to "Moderate") but never happened across groups (e.g. "Easy" to "Hard" or "Hard" to "Easy"). This indicated that although discrimination results differed from the prediction of PAM, this discrepancy seemed to be a minor one instead of a symbol of complete opposition.

To examine the discrepancy between predictions based on PAM and discrimination results statistically, we examined whether A-prime scores were different from each other within each PAM prediction level ("Easy", "Moderate", and "Hard" (see Table 2.11)). More specifically, repeated measures ANOVA were performed separately for each level---"Hard", "Moderate", and "Easy" to test whether any of the pairs was more difficult to discriminate than the others.

First we compared the A' scores of the seven pairs that PAM predicted to be easy to discriminate (i.e., [tʂ^{h}] vs. [ts], [ɕ] vs. [ʂ], [x] vs. [f], [i] vs. [y], [u] vs. [y], [ɑ] vs. [ə], [u] vs. [ɤ]) to examine whether the A' scores are consistent among the group, or whether any pairs were more difficult to discriminate than others. Repeated measures ANOVA analysis revealed that none of the mean A-prime scores of these contrasting pairs was found to be significantly different from each other (p =.72). This results indicated that all seven contrasting sound pairs are of the same difficulty level (easy) for the participants.

Similarly, repeated measures ANOVA analysis were conducted for five contrasting sound pairs that were categorized into "moderate" discrimination difficulty group by PAM (e.g. [k^{h}] vs. [k], [p^{h}] vs. [p], [ts^{h}] vs. [ts], [tɕ^{h}] vs. [tɕ] and [a] vs. [ɚ]).

The results also showed that no significant differences in A-prime scores of these five contrasts were found in this difficulty level (p =.73).

Discrimination difficulty "Hard" group based on PAM's prediction included three contrasting sound pairs---[t^h] vs. [t], [$tʂ^h$] vs. [tʂ] and [l] vs.[ɻ,]. Consistent with previous two groups, repeated measures ANOVA failed to reveal any significant differences in terms of A prime scores between these three test pairs (*p*=.96). This results confirmed that these three test pairs were of the same difficulty level to discriminate.

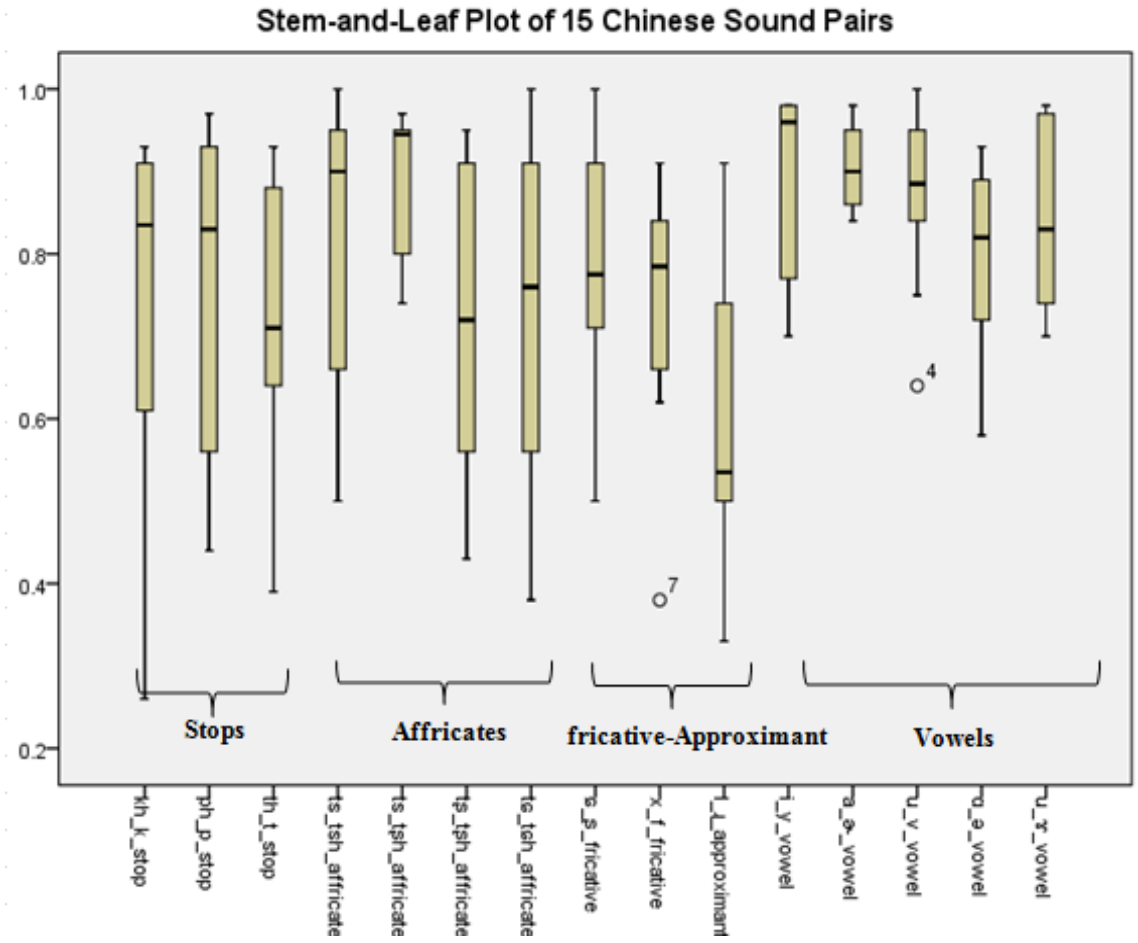
We ran these test in an attempt to verify visual observations; however, the results indicate that none of the contrasting pairs was found significantly different from other pairs within the same group of discrimination difficulty. These results may be due to actual lack of differences, or alternatively it may be due to the small sample size of this study.

In order to further explore the data, we also performed the repeated measures ANOVA on mean A prime scores to examine whether there was any difference in discrimination difficulty among consonant manner classes and vowels. Thus a test was performed separately for stop pairs ([k^h] vs. [k], [p^h] vs. [p], [t^h] vs. [t]), affricate pairs ([ts^h] vs. [ts], [$tʂ^h$] vs. [ts], [$tʂ^h$] vs. [tʂ], [$tɕ^h$] vs. [tɕ]), fricative-approximant group ([ɕ] vs. [ʂ], [x] vs. [f], and [l] vs. [ɻ]) and vowel groups ([i] vs. [y], [a] vs. [ɚ], [u] vs. [y], [ɑ] vs. [ə], [u] vs. [ɤ]). The stem-and-leaf plot (Figure 1) reports A prime scores separately for these groups.

The results of repeated measures ANOVA for stops showed no main effect of pair, indicating that none of A prime scores of these three stop pairs was significantly different from each other. The discriminability of Chinese stop pairs were of the same

level for the participants, although PAM predicted that [t^h] vs. [t] to be harder than the other two, and the mean A' score indeed showed a lower value for the pair (0.73) than the

Figure 2.1. Stem-and-Leaf plot of 15 Chinese sound pairs



other two pairs (0.80 and 0.77). It is highly possible that A prime scores for each subject is quite diversified, thus the slight difference in mean values is actually not statistically significant. As seen in Figure 1, although the mean A prime scores of [t^h] vs. [t] (indicated by the solid line in each bar) is much lower than the other two pairs, the range of distribution of each data point is very similar to that of the other two pairs. Accordingly, the results of both mean value and repeated measures ANOVA analysis

revealed that Japanese L2 learners had encountered roughly the same difficulty in discriminating Chinese stops.

The results of repeated measures ANOVA for affricates also showed that the main effect of pairs was not statistically significant, indicating that none of A prime scores of these three affricate pairs was significantly different from each other. Similar to the speculation made for Chinese stop pairs above, it is possible that the seemingly differences in mean values of A prime scores among four affricate pairs failed to accurately illustrate the whole picture of the distribution of each data points. As shown in Stem-and-Leaf plot in Figure 1, although the mean values of A prime scores for [ts^{h}] vs. [ts] and [tʂ^{h}] vs. [ts] pairs were very close to each other, the overall range of distribution for [ts^{h}] vs. [ts] actually was quite similar to the other two pairs ([tʂ^{h}] vs. [tʂ] and [tɕ^{h}] vs. [tɕ]), which had much lower mean values of A prime scores.

The results of repeated measures ANOVA for fricative--approximant group showed that none of A prime scores of these three pairs (two fricative pairs and one approximant pair) was found to be significantly different from each other. This indicated that Japanese L2 learners found that it was equally difficult to distinguish these Chinese fricative--approximant pairs. However, an exploratory pairwise comparison reveled that the difficulty in distinguishing approximant pair [l] vs. [ɻ] tended to be greater than that distinguishing fricative pair [ɕ] vs. [ʂ] (p = .043). Similarly, the difficulty in distinguishing approximant pair [l] vs. [ɻ] was also found to show the tendency to be greater than that of fricative pair [x] vs. [f] (p = .035). However, both tendency failed to reach statistical significance.

Finally, the results of repeated measures ANOVA for the vowel group showed that none of A prime scores of these five vowel pairs was found to be significantly different from each other. Exploratory pairwise comparison of each vowel pair did not reveal any significant difference either. This indicated that Japanese L2 learners found that it was equally difficult to distinguish these Chinese vowel pairs. This finding is generally consistent with the results in our categorization based on mean A prime values except for one vowel pair ([ɑ] vs. [ə]). As indicated in Table 2.11, the A prime scores were found to rank at the upper level of the group for four vowel pairs [i] vs. [y], [a] vs. [ɚ], [u] vs. [y], [u] vs. [ɤ], while [ɑ] vs. [ə]'s A prime score was found to be lie in the middle of the group. Although the mean value of A prime scores for [ɑ] vs. [ə] pair is the lowest among five vowel pairs, the statistical analysis seemed to suggest that although the mean of this pair was lower than other vowel pairs, focusing on mean difference alone actually failed to capture the general distribution pattern. As confirmed by the stem-and-leaf plot in Figure 1, the range of distribution of A prime scores for [ɑ] vs. [ə] pair did not seem to differ greatly from that of other groups.

2.4. General Discussion

In this study, we first conducted categorical mapping experiment to investigate how Mandarin Chinese vowels and consonants were perceived by Japanese native speakers in terms of the most resembling Japanese categories and provided a quantitative measure of this perceived phonetic distance for each Chinese sound and its identified Japanese category using fit index. The results indicated that most Chinese sounds were categorizable into Japanese native categories with a few exceptions. For consonants, Chinese retroflex affricate pairs ([tʂ^{h}]-[tʂ]) and fricative [x] were uncategorizable into one

single Japanese category since they all fell in between two different categories. For Chinese monophthong vowels, Japanese native speakers also failed to reach a consensus on which Japanese vowel category should back vowel [ɑ], rounded vowel [y] and mid back vowel [ɤ] be classified into. Chinese diphthongs were found to be especially hard for Japanese native speakers to categorize. Out of six diphthongs examined, only half of them ([ɥe], [wo] and [ei]) were categorized into Japanese category with relatively low fit indices (0.75-2.97). For the other three diphthongs ([ou], [ai], and [ɑu]), Japanese listeners were uncertain whether they should be categorized into one vowel category or two vowels. For instance, some Japanese listeners heard Chinese [ai] as a Japanese two vowel combination [a]+[i], while other heard it as one single Japanese vowel [e]. These data seemed to suggest that categorizability of Chinese sounds was determined by whether there exists a counterpart category with the same feature, place and manner of articulation in Japanese phonemic inventory. Compared to Chinese, Japanese lacks the category of retroflex ([tʂ^{h}]-[tʂ]), rounded vowel ([y]) and mid back vowel ([ɤ]), which were exactly the sounds the Japanese listeners had difficulty in categorizing. However, successfully classification of other Chinese sounds (e.g., retroflexed vowel [ɚ], retroflex fricative [ʂ]) indicated that the categorizability was not only related to the existence of a counterpart category but also whether this sound is perceived to be in proximity to more than one categories in their native language.

Among sounds that were categorizable, it was found that certain Chinese sounds were found easier to categorize into Japanese native categories than others. As indicated in Table 2.6 and 2.7, we can see that in general, sonorants (nasals [ŋ, m, n] and approximants [l, ɻ]) were the easiest to categorize, followed by obstruents (stops,

affricates and fricatives), vowels especially diphthongs posed the greatest difficulty for classification for Japanese native speakers.

The cross-language mapping data also revealed some interesting findings about how Chinese aspirated and unaspirated stops and affricates were perceived by Japanese native speakers who distinguish these sound categories based on voicing in their native language. As indicated in the results of category mapping experiment, aspirated and unaspirated pairs of Chinese stops and affricates were the most frequently classified into the same Japanese voiceless category. For instance, stop pair [k^h] and [k] were both classified into Japanese voiceless stop [k] and affricate pair [ts] and [ts^h] were classified into Japanese voiceless affricate [ts]. This finding was in general consistent with acoustic comparison of VOT durations of Japanese and Chinese stops in previous studies, indicating that the use of acoustic cues by listeners in categorical mapping tasks. It is claimed that VOTs of Japanese voiceless stops fall in between the two general groupings of voiceless stops: short lag (0-25ms for unaspirated stops) and long lag (60-100ms for aspirated stops) in world's languages (Riney, et al, 2007). In a cross-language study of voicing contrasts of stops conducted by Shimizu (Shimizu, 1989) the mean VOTs and ranges for Japanese voiceless stops [p], [t], and [k] and voiced stops [b], [d], and [k] were acoustically measured and reported in Table 2.12. Chao & Chen also acoustically measured the VOTs of Chinese aspirated and unaspirated stops ([p], [p^h], [t], [t^h], [k], [k^h]) and the values were reported in Table 2.13 (Chao & Chen, 2008).

A rough comparison between Chinese and Japanese VOT values above revealed that the VOT ranges of Japanese voiceless stops overlap with the VOT ranges of both Chinese aspirated and unaspirated stops, while the VOT ranges of Japanese voiced stops

Table 2.12. VOT ranges and general means (in ms) for Japanese stops

	p	**t**	**k**	**b**	**d**	**g**
Min	15	15	45	-45	-10	-20
Max	60	90	100	-92	-70	-105
General Means	44	27	68	-72	-58	-64

Table 2.13. VOT ranges and general means (in ms) for Mandarin stops

	p^h	**t^h**	**k^h**	**p**	**t**	**k**
Min	35	45	50	7	7	15
Max	147	123	138	65	33	65
General Means	82	81	92	14	16	27

do not overlap with either Chinese stop categories. Thus we can see that classification of Chinese sounds into Japanese categories were based on the perception of phonetic features (VOT in this case), which readily explained why both Chinese aspirated and unaspirated stops are both assimilated into Japanese voiceless stop categories by Japanese listeners. In addition, in general the fit indices of Chinese aspirated stops and affricates for the identified Japanese voiceless category were higher than their unaspirated counterparts except for affricate pair [ts]-[ts^h], since the fit index score of [ts] is slightly higher than that of [ts^h] (3.46 vs.3.01) for the identified Japanese category [ts]. This results indicated that Chinese aspirated stops and affricates were perceived by Japanese native speakers as better exemplars of the voiceless category they had identified than their unaspirated counterparts, assumedly due to the closeness in VOT ranges between Chinese aspirated consonants and Japanese voiceless consonants.

This Chinese-Japanese category mapping data as well as perceived phonetic distance for each mapping measured in fit index enabled us to generate a set of 15

testable sound contrasts with predicted discrimination difficulties based on PAM theoretical framework. In order to find whether these predictions are valid we further conducted discrimination experiment which tested the actual discriminability for each contrast pair judged by Japanese L2 learners with limited Chinese learning experience. The comparison between predicted discrimination difficulties and their A prime score ranking among 15 sound contrasts revealed that the discriminability between sound contrasts depended on the perceived phonetic proximity in general, although not all the predictions based on PAM were supported. This discrepancy implied that PAM framework cannot be readily applied to the context of L2 learning without modification.

Now we take a closer look at these four contrast pairs out of fifteen contrasts that were found to have different discriminabilty than predicted. The discrepancy occurred for two sound contrasts of CG assimilation type---[a] vs. [ɚ] and [ts^{h}] vs. [ts], in which PAM predicted the discrimination between them to be "moderate", while it was found that they were actually very easy to discriminate. In both cases, the discrimination difficulty was overestimated by PAM. One possible explanation was that although perceived phonetic distances between these two sound contrasts were not small, the distinctiveness of the phonetic features in one of the sound pairs might have helped with the discrimination task. It is likely that the aspiration feature of affricate [ts^{h}] in [ts^{h}] vs. [ts] pair and rhotacization feature of vowel [ɚ] in contrast pair [a] vs. [ɚ] might be perceptually salient so that it helped decrease the discrimination difficulty. For the other two fricative pairs that were found to be harder to distinguish while PAM predicted easy discriminability, contrast pair [ɕ] vs. [ʂ] also belonged to CG type while [x] vs. [f] belonged to UC type. For fricative contrast [x] vs. [f], it is important to note that [x] is not categorizable to any

single Japanese consonant category because it was heard as both Japanese [h] and [k], while [f] was categorized as a poor exemplar of [h]. The fit indices for [x] vs. [f] classified as Japanese category [h] were also close to each other (1.98 vs. 1.29). Accordingly, this overlap in assimilation category might have narrowed perceived phonetic distance between [x] vs. [f] and added extra difficulty in distinguishing these two sounds. For [ɐ] vs. [ʂ], however, we do not yet have an explanation on why the same retroflex feature of fricative [ʂ] behaved in the opposite way than vowel pair [a] vs. [ɚ]. Perhaps the retroflex feature in fricatives is perceptually much more difficult than retroflex feature in vowels, after all, it is fairly rare to have retroflexed vowels. The purpose of this study is not to prove the impeccability of PAM by trying to show that the findings in this study were exactly the same as PAM's predictions. The inconsistency here suggested that revision of PAM framework is necessary in order to apply it to L2 learning. One revision is suggested for predictions on discriminability of GC assimilation type sound contrast, since the phonetic features possessed by one sound in the contrasts might perceptually facilitate or hinder discrimination. Another revision is suggested for UC Type sound contrast, because the shared assimilation category between categorizable sound and uncategorizable sound might exert influence on perceived phonetic distances between them.

2.5. Conclusion

The results obtained in this study provided important insight on how Japanese native speakers perceive Chinese sounds in terms of their native category, and indicated that PAM cannot be extended to L2 learning, especially for the prediction of certain UC and CG type discrimination. This finding is consistent with previous research. As

reviewed in Introduction session, Guion et al (2000) testified predictions based on PAM by investigating Japanese native listeners' perceptual mapping of a limited number of English consonants onto Japanese categories. Similar to the findings in current study, it was found that out of three contrasts, the discriminability of [s]-[θ] contrast which fell into PAM's UC assimilation type was not consistent with PAM's predictions. The poor prediction made by PAM especially on UC type that has overlap in assimilated category suggested that the prediction of discriminability between uncategorizable and categorizable L2 sounds should also depend on relative goodness rating to the assimilated category. In addition, it is important to note that the notion of perceived phonetic distances between L2 sounds and the native category proposed by PAM framework was measured by conducting perceptual mapping experiment on subjects who had no learning experience of L2 language. Since categorical discrimination experiment in this Chapter tested L2 learners with limited language experience, some of the discrepancy is also likely to result from possible changes in the perceived phonetic distances of L2 sounds due to language learning experiences, especially for the ones turned out to be easy although PAM predicted "difficult". Future researches are necessary to explore whether language experiences can alter the perceived phonetic distances between L2 sound and the classified L1 category. If this is the case, we want to know whether the learning effects are the same for sound pairs with various predicted discrimination difficulties. We are also interested in researching on whether some distinctive phonetic features such as aspiration and rhotacization will effect discrimination in a positive or negative way in future study.

The findings of this study in both categorical mapping and discrimination experiments also provided important pedagogical implications that can serve as teaching guide for language instructors teaching Mandarin Chinese to Japanese-speaking L2 learners. Since the results above have confirmed that certain Chinese sound pairs were harder to discern than others, it is suggested that language instructors put more efforts into these "hard" to discriminate pairs. For instance, instructors can create specific excises or activities that specifically focus on these pairs to help students first perceptually discriminate and produce L2 sounds more accurately later.

Chapter III

TRAINING JAPANESE L2 LEARNERS' TO IDENTIFY MANDARIN CHINESE CONSONANT CONTRASTS

3.1. Introduction

As discussed in Introduction chapter, human beings' perception of speech sounds is shaped by many years of extensive exposure to our native language. During this process, we gradually lost sensitivity to acoustic cues that are not phonemic in our native language (L1) and become only attentive to those language-specific distinctive contrasts after these cues had been strengthened repetitively with linguistic exposure over life-span. This has posed great difficulty for learners of a second language (L2) to perceive and further produce nonnative sounds. Previous researches have demonstrated that native speakers outperform L2 learners in discrimination of certain sound contrasts (Miyawaki et al, 1975; Werker and Logan, 1985). Nevertheless, individual perceptual system is not static and L2 learners' ability to distinguish non-native sounds can be improved with increased experiences in the target language, even by intensive laboratory training in a short period of time.

This plasticity of human speech perceptual system was extensively confirmed by previous studies. Simply using synthetic stimulus familiarization and training with immediate feedback techniques, McClaskey, Pisoni, & Carrell (1983) successfully trained English native speakers who only distinguish between voiced and voiceless in their L1 stops to identify an additional voicing category (pre-voiced stop), and the training effects were also successfully transferred to novel stimuli produced with a different place of articulation (from labial to alveolar).

Another classic example in cross-linguistic speech perceptual training studies is to train Japanese L2 learners of English to perceive English liquids [r] and [l]. Japanese speakers struggle in discrimination between these two sounds even after living in a English-speaking country for years (MacKain, Best, & Strange, 1981), because this sound pair was phonemically distinctive in English but not in Japanese (Goto, 1971). Japanese only has a tap or flap phoneme [ɾ], which has a diverse variation of realization: [ɹ, ɻ, ɖ, ɭ, r, l] (Best & Strange, 1992; Ingram & Park, 1998; Kochetov & Smith, 2009; Okada, 2005). Strange & Dittmann (1984) made the first attempt to modify Japanese listeners' perceptual system using laboratory-based intensive training. The performance of eight adult female Japanese native speakers on discrimination task on synthetic sound contrast "rock" and "lock" slightly improved after being trained on AX discrimination task with immediate feedback for about three weeks. In this study, synthetic stimuli with very little variability were used. Although the training effects were successfully transferred to different tasks (identification and oddity discrimination) and novel stimuli ("rake"-"lake"), yet it failed to extend to generalization tests on natural speech words.

In order to improve limited effectiveness of laboratory training on Japanese listeners to identify English [r] and [l] in previous study, Logan et al (1991) made several critical improvements of the training paradigm employed in Strange & Dittmann's (1984) study. The most important change was using natural speech stimuli produced by multiple (five) talkers in different phonetic environments rather than using synthesized stimuli with little variability in training sessions. The motivation of this change was inspired by the results of a well-known psychology study (Posner & Keele, 1968) which found the group trained on stimuli with high degree of variability outperformed group trained with

low degree of variability in classifying a visual stimuli task. Thus it was believed that variability in stimuli could provide full range of acoustic cues for category characterization and further category formation. In addition, instead of using AX discrimination tasks to train and test listeners, two-alternative forced-choice identification tasks were employed since AX training procedure was thought to fail to direct listeners attention to categorization by simply focusing on low-level, sensory-based information in stimuli presented, thus would be less likely to generate training effects that are transferable to other contexts. The results of this study confirmed the effectiveness of training by successfully generating robust improvement on Japanese listeners' performance on identifying English [r] and [l] pair. In this study, the training paradigm which emphasizes on stimulus variability by using natural speech tokens produced by multiple talkers in different phonetic environments was referred to as HVPT (High Variability Phonetic Training). In a later study, Lively et al (1993) also confirmed the importance of stimulus variability in phonetic training experiments, especially talker variability, by comparing the performance of two groups of Japanese listeners. The first group was trained using identification task with [r]-[l] produced by multiple talkers in three different syllable positions, while the second group was trained with the same consonant contrast produced by only one talker in five different syllable positions. Both groups showed improved performance in posttest compared to pretest, although the performance improvement in the first group also generalized to new words spoken by a familiar talker and new words by a new talker, while it failed to generalized to tokens spoken by a new talker for the second group. In a follow-up study, Bradlow, Akahane-Yamada, Pisoni, & Tohkura (1999) not only replicated the efficacy of HVPT in training

Japanese listeners to identify English [r]-[l] but proved successful retention of the training effects even after 3 months.

HVPT procedure has been extensively used in L2 perceptual training study and was proved successful not only in training Japanese L2 learners to perceive English [r]-[l] contrast (Lively et al., 1994; Iverson et al, 2005; (Bradlow, 2008; Bradlow, Pisoni, Akahane-Yamada, & Tohkura, 1997), but for cross-language perceptual learning in other language pairs. Hirata et al (2007) successfully trained native English speakers to identify Japanese long vowel and short vowel using three types of training which differed in sentential speaking rates---slow-only, fast-only and slow-fast. Wang et al (1999)'s study proved HVPT paradigm is not only effective in modifying listeners' speech perceptual system in the segmental domain but in suprasegmental domain. After been trained with Mandarin Chinese natural words produced in various phonetic contexts by multiple talkers, American learners of Mandarin not only improved their performance in identifying four Chinese tones tasks but showed robust retention in a six-month retention test afterwards.

Wong (2012) also compared HVPT with LVPT (Low Variability Phonetic Training) in training Cantonese native speakers to identify and produce English vowel [e]-[æ] contrast that was absent in Cantonese vowel inventory and confirmed the superiority of HVPT over LVPT in terms of training effects. Twenty-two Cantonese L2 learners of English were trained using HVPT paradigm, in which they were asked to perform two-alternative forced choice identification task of 60 stimuli produced by six English native speakers. In contrast, nineteen Cantonese L2 learners received LVPT training and the only difference was that the stimuli they received the training were

produced by only one female native speakers of English. The results indicated that although all subjects showed signification perceptual learning and generalization to new words and new speakers after the training, subjects in HVPT group showed higher degree of improvement and more robust transfer from perception to production as compared with LVPT group.

Although the effectiveness of HVPT has been widely confirmed, it is important to note that stimulus variability probably is not the only effective factor that researchers could possibly manipulate in laboratory training to alter L2 learners perception. Iverson, Hazan, & Bannister (2005ue) conducted a perceptual training study to investigate the relative effectiveness of four training techniques: HVPT, All Enhanced, Perpetual Fading and Secondary Cue Variability in training Japanese L2 learners identify English [r]-[l] contrast. These three techniques other than HVPT manipulated F3 or F2 variability in stimuli by signal processing techniques. The findings indicated that all four training paradigms were effective in helping improve Japanese listeners' performance on identification of English [r]-[l] sound contrast tasks, and no differences were found between these techniques.

Previous training studies reviewed above mainly focused on English as the target language, although there are a few studies looked at Mandarin Chinese by training English native speakers to perceive Chinese tones (Wang et al, 1999, 2003). In this chapter, we are going present a perceptual training study on Japanese L2 learners trying to distinguish Mandarin Chinese consonant pairs using HVPT paradigm. In previous chapter we discussed the pedagogical implication of having the knowledge of perceived phonetic distances of Chinese sounds to their classified Japanese categories. One

objective of this training study is to examine whether intensive laboratory training could facilitate Japanese L2 learners in improving their performance on identifying Chinese sound contrasts. If so, how the improvement differ with regard to sound pairs with different phonetic distances. This finding could inform L2 instructors so that they could develop effective teaching strategies based on this information to tackle specific sound pairs and accelerate learning in L2 classroom. In order to do this, we selected two Chinese consonant pairs, one from "hard" discriminability group ([t] vs. [t^h]) and the other from the "easy" group ([ts] vs. [tʂ]) respectively as the target sound contrasts based on the results from categorical mapping study in chapter 2.

In addition, the results of categorical mapping and discrimination draw our attention to two phonetic features specific to Mandarin Chinese: aspiration and retroflexion. Especially for Chinese retroflex, the discrepancy between PAM's prediction and discrimination results for pair [ɕ] vs. [ʂ] made us speculate the phonetic feature of retroflexion might also have influence on Japanese L2 learners' perceptual categorization in addition to perceived phonetic distances based on which PAM's predictions were laid out. Selection of these two pairs is also motivated by another research interest: to explore whether these two phonetic features are learnable for Japanese L2 learners whose L1 lacks distinction based on either aspiration or retroflexion. Specifically, whether Japanese L2 learners' perception can be altered by language experiences, in this case, intensive laboratory-based training. Accordingly, [t] vs. [t^h] and [ts] vs. [tʂ] are perfect sound contrast candidates since stop contrast [t] and [t^h] only differ in terms of aspiration, while affricate contrast [ts] and retroflex [tʂ] only differ in terms of retroflexion.

To sum up, this chapter presents a speech perceptual training study on an understudied language pair---Chinese-Japanese by training Japanese L2 learners to distinguish Chinese consonant pairs using HVPT. The objective of this experiment is threefold: (1) to investigate whether HVPT paradigm is effective in altering Japanese listeners' perception of Chinese consonant pairs after a short period of intensive laboratory training. If so, whether the improvement in performance can be extended to novel stimuli produced in different phonetic environments by different talkers. In order to do this, we employed a classic pretest-posttest test design followed by two generalization tests (G1 and G2), in which G1 tested listeners on identification of target stimuli produced in different vowel environment by a familiar talker, while G2 used stimuli produced in different vowel environments by a unfamiliar talker. (2) if training effects is confirmed, whether the degree of perceptual learning due to laboratory training is the same for sound contrasts of different discriminability levels predicted by PAM (3) to investigate whether HVPT is effective in modifying Japanese L2 learners' perception of two different phonetic features specific to Chinese phonological categories---aspiration and retroflexion. In order to accomplish the last two goals, we have selected two sound contrasts---[t] vs. [t^h] (distinguished by aspiration) and [ts] vs. [tʂ] (distinguished by retroflexion) to respectively represent " hard" and " easy " discriminability contrasts.

3.2. Methods

3.2.1. Participants

Nine native speakers of Japanese (6 female, 3 male, mean age= 22.2), who had varying experience of learning Chinese (2.5-34 months), recruited at University of Oregon, participated in the training study (see Table 3.1). Out of all Japanese native

speakers, four were attending Chinese beginner level course (CHN 101) at the time of data collection, and the rest had learned Chinese before but were not taking any Chinese courses at the time of experiment. All participants came from areas where standard Tokyo dialect is spoken except one came from western (Kansai) area of Japan. The majority of Japanese L2 learners never went to China, except one participant went to Hong Kong or Taiwan each year for one month and another participant stayed in China for one week for sightseeing. All reported daily use of Japanese. The participants were randomly assigned to one of the following group: training group (n = 5, mean age= 20.4) or control group (n =4, mean age= 19.8). Accordingly, five Japanese native speakers participated in training group and four native speakers participated in control group. None of participants reported hearing problem and they were all paid for their participation in this training experiment.

Table 3.1. Language background of Japanese L2 participants

	Training (n = 5)	*Control (n =4)*
Age (years)	20.4 (19-21)	24.5 (21-27)
AOL (years)	19.8 (19-21)	20.3 (18-25)
Chinese instruction (months)	4.7 (2.5-12)	15 (3-34)
Length of stay in China (months)	0 (0-0)	2.0 (0-8)

3.2.2. Stimuli

Six Mandarin Chinese native speakers (3 male, 3 female, mean age=20.6) from mainland China produced Chinese speech stimuli for this training experiment. All participants were attending the University of Oregon at the time of data collection. Their average length of stay in the US was 2.7 years (range: 1.5 years – 3.5 years). All

Mandarin Chinese native speaker were from northern China. The dialects they speak all belong to the northern dialect family of Chinese, which is very similar to Mandarin Chinese. All reported daily use of Mandarin Chinese.

As discussed above in Introduction, the sound pairs used for this training experiment were two consonant contrasts [t] vs. [t^h] and [ts] vs. [tʂ] selected as representative sound pairs from "hard" and "easy" learning difficulty levels based on the findings of categorical mapping and discrimination study in Chapter 2. Six native Mandarin Chinese speakers (3 male and 3 female) coded as M1, M2, M3, F1, F2 and F3 produced test consonant stimuli in a range of allowable monosyllabic words embedded in a carrier sentence "*qing du___san bian" (Please read ___ three times)*. For instance, tokens of consonant stimuli [t^h] were obtained by embedding monosyllables [t^ha], [t^han], and [t^hai] in the carrier sentence "*qing du* [t^ha]/[t^han]/[t^hai] *san bian" (Please read* [t^ha]/[t^han]/[t^hai] *three times)*. Before recording, Chinese native speakers were provided with a list of syllables containing target consonant pairs written in Chinese orthography with pinyin and tone annotation (the list of the Chinese characters used for stimuli elicitation was presented in Appendix B). All the target consonant pairs were produced in the same vowel environment with the same tone twice, and the second production was selected so as to eliminate possible effluences introduced by different phonetic environment. For instance, contrast pair [t^h] vs. [t] were both produced in an open syllable [t^ha] and [ta] followed by the same vowel with the first tone. In the same vein, contrast pair [ts] vs. [tʂ] were produced in syllable [tsai] and [tʂai] followed by diphthong [ai] with the same third tone.

This training included consecutive five days of training on two target consonant contrasts [t^h] vs. [t] and [ts] vs. [tʂ] and four types of tests: pretest, posttest and two generalization tests (Gen1 and Gen2). Pretest and posttest contained the same set of 64 stimuli [8 syllables× 2 consonants× 2 repetitions× 2 contrasts], while training sections on each day contained 120 stimuli [15 syllables × 2 consonants × 2 repetitions × 2 contrasts]. These 120 stimuli included all the 64 stimuli used in pretest/posttest section, and also included additional 56 stimuli of seven new syllables to further diversify syllable environments of target consonants. The first generalization test (Gen1) and the second generalization test (Gen2) contained 80 novel stimuli of ten syllables that never appeared in previous sections [10 syllables × 2 consonants × 2 repetitions × 2 contrasts]. Gen1 consisted of novel stimuli produced by a familiar speaker (F2 who produced the stimuli that were used on the third day of training) that never appeared in either pretest or training sessions. In contrast, Gen2 consisted of stimuli produced by a novel speaker (F3 whose production was not used in either pretest/posttest or the training session). The detailed stimuli list can be found in Appendix B. The motivation of adding two additional tests (Gen1 and Gen 2) is to investigate whether the training effects, if any, could be generalized to tokens that Japanese listeners were not exposed to. The speakers were recorded individually in a sound-attenuated booth using a flash digital recorder (Marantz PMD 670) and a standing microphone (SHURE Beta 87) at a sampling rate of 44 kHz and 16-bit quantization. All the consonant stimuli were extracted from carrier sentences and normalized to 70dB.

3.2.3. Procedure

This training experiment consisted of four sections: pretest, posttest and two generalization tests. The effects of training were accessed by comparing identification accuracy in pretest and post-test administered before and after a 5-day training period. Generalization of training effects to novel stimuli spoken by a familiar talker and novel tokens spoken by new speakers were accessed by comparing the performance of pretest and two generalization tests.

On the first day of the experiment, subjects in training group participated in pretest followed by a training session. From day 2 to day 4, subjects repeated the training procedure with stimuli produced by different native Chinese speakers on each day. On the last (fifth) day of training, after completing the same training session as previous 4 days, subjects were asked to participate in posttest and two generalization tests (Gen1 and Gen2). A control group of four subjects who only took the pretest on the first day and posttest on last day after the same five day interval were also included so as to guarantee that any training effects obtained by comparing pretest with posttest were not due to simply repeating the test twice. The detailed experiment procedure were listed in Table 3.2.

Table 3.2. Overview of the experiment procedure for the Control and Training groups

Session (day)	**Procedure**	**Control**	**Training**	**Speaker**	**Carrier Sentence**	**Number of Trials**
1	Pretest	√	√	M1	*"qing du___san bian" (Please read ___ three times*	8*2*2*2=64
	Training1		√	F1		15*2*2*2=120
2	Training2		√	M2		15*2*2*2=120
3	Training3		√	F2		15*2*2*2=120
4	Training4		√	M1		15*2*2*2=120
5	Training5		√	M3		15*2*2*2=120
	Posttest	√	√	M1		8*2*2*2=64
	Gen1	√	√	F2		10*2*2*2=80
	Gen2	√	√	F3		10*2*2*2=80

In this experiment, all speech stimuli were presented to participants at a comfortable listening level over headphone in a sound-attenuated booth using E-prime software. Each test or training session consisted of two blocks for two different contrasts. The order of stimuli presented in each block were randomized. The order of presenting two target consonant pairs to be tested in each block were counterbalanced across participants. In the first block, Japanese listeners were instructed to focus on the consonant portion of each sound and identify whether the consonant they had heard were [t^h] or [t] by pressing key 1 or 2. Similarly in the second block, they were asked to identify whether the consonants they had heard were [ts] or [tʂ] by pressing key 1 or 2. In all training sections, feedback was given after each trial. The correct answers were shown on the next screen after the subjects submitted their answers. In all test sections (pretest, posttest, Gen1 and Gen2), however, no feedback was given regarding whether they had made the correct choices.

3.2.4. Analysis

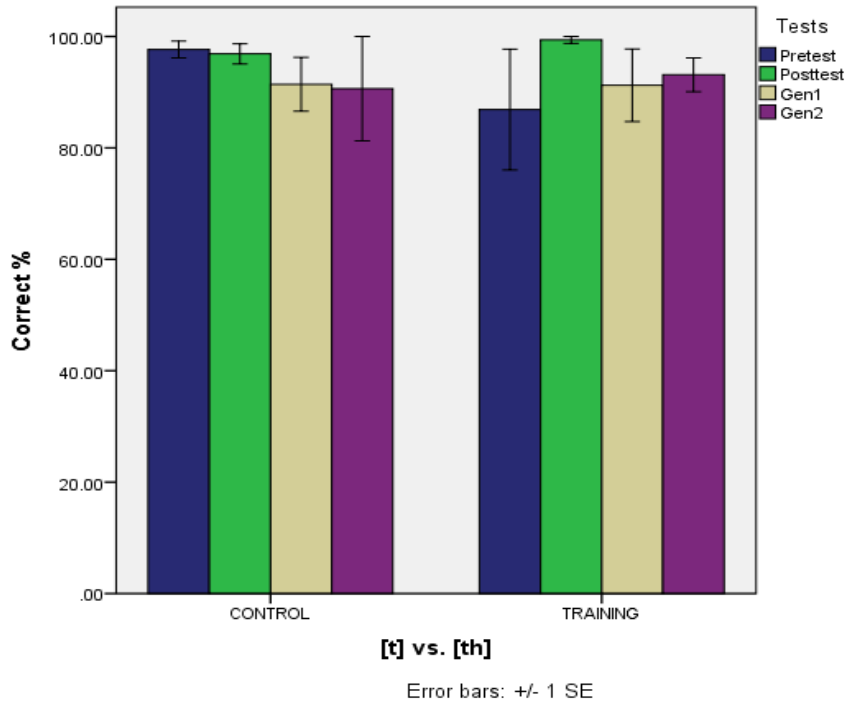
The percentage of correct responses was calculated for each Japanese listener. This accuracy score was used as dependent variable in the following analysis. In order to investigate the effect of training on overall performance in identification accuracy, the accuracy scores were submitted to a three-way ANOVA with Group (training, control) as the between-subject factor, and test (pretest, post-test) and contrasts ([t] vs.[t^h] and [ts] vs. [tʂ]) as the within-subject factors.

3.2.5. Results

Figure 3.1 and 3.2 present descriptive statistics of the overall performance of Japanese listeners in control and training group at four different tests using the mean

percentages of correct responses in identification tasks for two different contrasting pairs

Figure 3.1. Identification mean accuracy scores (in percentages) of Control and Training group at four tests (Pretest, Posttest, Gen1, and Gen2) for contrast pair [t] vs.[t^h]. Error bars indicate +/- *SE*.



[t] vs.[t^h] and [ts] vs. [tʂ]. In general, the accuracy percentages of Japanese listeners in training group tended to outperform that in control group. In order to examine the effect of training statistically, we submitted the accuracy percentages of each Japanese listener to mixed ANOVA analysis and the results were summarized in Table 3.3. The analysis yielded significant main effects of Test [$F(3, 21)=4.14$, $p=.019$] and Contrast [$F(1, 7)=9.57$, $p=.017$]. This indicated that the overall identification accuracy scores were significantly different across pretest (80.51%), posttest (90.04%), Gen1 (88.52%) and Gen2 (89.96%) tests regardless of different group and contrasting pairs. In the same vein, if we ignore three Tests types (pretest, post, Gen1, and Gen2) and Group type (training or

control group), the identification accuracy scores were higher for [t] vs. [th] pair (93.55%) than [ts] vs. [tʂ] pair (80.96%).

Figure 3.2. Identification mean accuracy scores (in percentages) of Control and Training group at four tests (Pretest, Posttest, Gen1, and Gen2) for contrast pair [ts] vs. [tʂ]. Error bars indicate +/- *SE*.

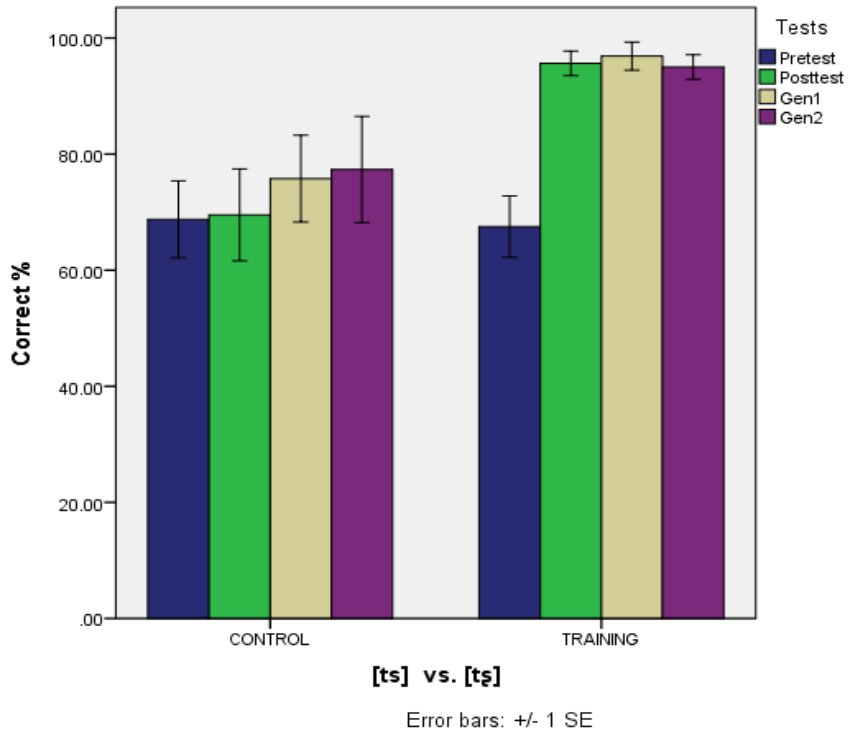


Table 3.3. Summary of significant main effects and interaction in 4x2x2 ANOVA

Effect	***F***	***df***	***p***
Test	4.14	(3, 21)	0.019
Contrast	9.57	(1, 7)	0.017
Test x Contrast	4.38	(3, 21)	0.015
Group x Test	3.82	(3, 21)	0.025
Group x Contrast	4.56	(1, 7)	0.070
Group x Test x Contrast	0.74	(3, 21)	0.538

In addition, three-way ANOVA analysis revealed significant interaction between Group and Test [$F(3, 21) = 3.82$, $p = .025$] Test and Contrast [$F(3, 21) = 4.38$, $p = .015$] but

not between Group and Contrast [$F(1, 7)=4.56$, $p=0.07$] (Table 3.3). For interaction between Group and Test, the means of accuracy scores for two groups across four tests were displayed in Figure 3.3, and CT and TR represented control and training groups respectively. The result confirmed the overall effectiveness of training since there was

Figure 3.3. Mean accuracy scores for control and training group across four tests

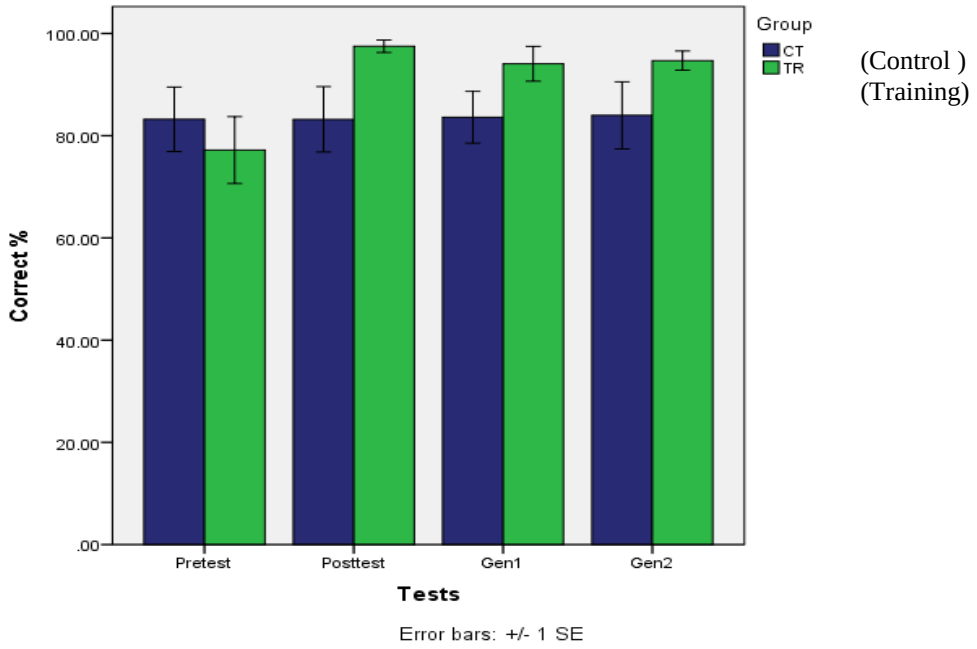
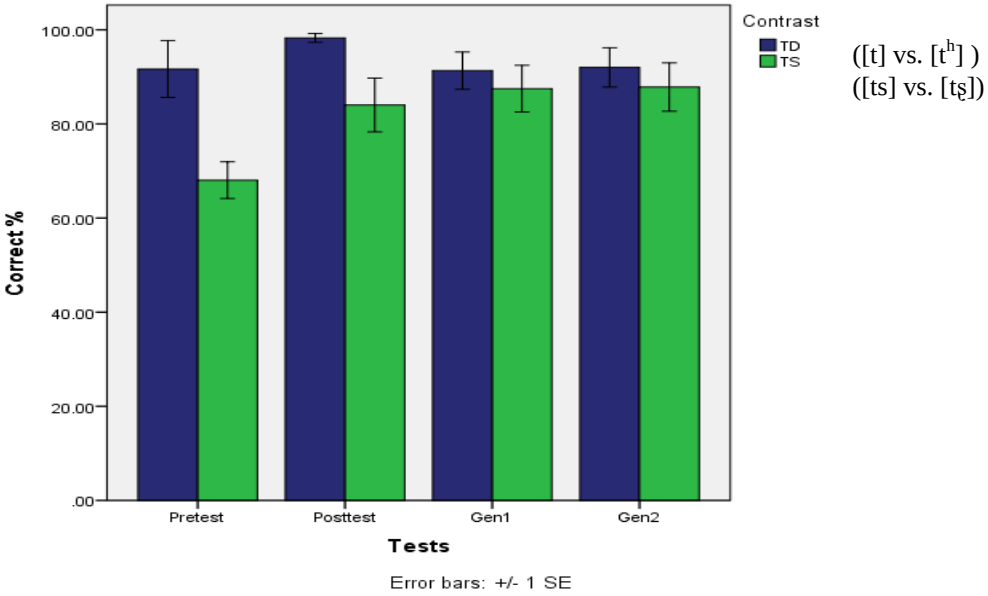


Figure 3.4. Mean accuracy scores for [t] vs. [t^{h}] and [ts] vs. [tʂ] contrasts across four tests



significant improvement of performance over four tests (pretest, posttest, Gen1 and Gen2) but the amount of improvement differed for control and training group. Alternatively, we can also say that training group outperformed control group in all four tests, yet the accuracy differences differed across tests. This Group x Test significant interaction was further examined in a follow-up simple effects tests of four different tests in both control and training groups (collapsed across two contrasts). The tests revealed that none of the accuracy improvement for control group between pretest and three other tests were significant [p>.05], while all accuracy improvement in training group were found significant (pretest-posttest [p=.001], pretest-Gen1 [p=.000], pretest-Gen2 [p=.020]). This would be fairly straightforward if we take a look at the increases in mean accuracy scores for each group. The identification accuracy for training group increased 19.063% from pretest to posttest, while decreased 0.001% for control group, probably because Japanese learners who did not receive training simply provided their responses in either pretest or posttest by random guessing. Similarly, training group improved their accuracy score of by 15.63% from pretest to Gen1, while control group barely made any improvement (0.39%). In addition, the accuracy score for training group increased 18.13% from pretest to Gen2, while the extremely limited increase in accuracy scores by control group (0.78%) is probably more accurate to be seen as chance variation than any systematic improvement. This results can be interpreted that the performance of Japanese listeners in control group did not improve significantly across four tests after simple repetition of the test stimuli, while Japanese listeners who participated in training session improved significantly in terms of identification accuracy after the training. Similarly, simple effects tests of control and training groups in four different tests revealed that the

accuracy difference between control and training groups (collapsed across two contrasts) was significant in only posttest [$F(1,7)=9.87$, $p=.016$] but not in pretest [$F(1,7)=.910$, $p=.372$], Gen1 [$F(1,7)=2.809$, $p=.138$] and Gen2 [$F(1,7)=2.051$, $p=.195$]. This indicated that the performance of Japanese listeners in identifying Chinese sound pairs in control group and training group did not differ at the onset of the experiment. After 5-day consecutive training period elapsed, Japanese listeners who participated in training session did perform significantly better in posttests than those who did not. However, their performances in two generalization tests were not superior to listeners who did not receive training, indicating that training effects failed to generalize when Japanese listeners were tested on novel speech stimuli either produced by a familiar talker or a new talker.

Significant interaction was also found between Test and Contrast [$F(3, 21)=4.38$, $p=.015$] (Table 3.3). The means of accuracy percentages for two contrasts across four tests were displayed in Figure 3.4, in which TD and TS represented [t] vs. [t^h] and [ts] vs. [tʂ] contrast pairs respectively. The results confirmed the overall effectiveness of training since there was significant improvement of performance over four tests (pretest, posttest, Gen1 and Gen2) but the amount of improvement differed for [t] vs. [t^h] and [ts] vs. [tʂ] contrasts. This significant interaction was further examined by a follow-up simple effects tests of four different tests for both contrasts (collapsed across two groups). The tests revealed that none of the performance improvement between pretest and other three other tests were significant [$p>.05$] for [t] vs. [t^h] contrast, while the accuracy improvement between pretest and three other tests for [ts] vs. [tʂ] contrast were all found significant (posttest-pretest [$p=.005$], Gen1-pretest [$p=.004$], Gen2-pretest [$p=.009$]). The

improvement in mean accuracy score will provide a detailed picture of the training effects for each sound contrast. For [ts] vs. [tʂ] contrast, Japanese listeners improved their identification accuracy by 14.45% from pretest to posttest, while the improvement for [t] vs. [t^h] contrast is merely 4.61%. The accuracy scores for Japanese listeners increased by 18.20% from Gen1 to pretest for identifying [ts] vs. [tʂ] contrast, while decreased by 2.19% for [t] vs. [t^h] contrast. In addition, the improvement for Japanese listeners from pretest to Gen2 was as large as 18.67% for [ts] vs. [tʂ] contrast, while the improvement for [t] vs. [t^h] contrast was found very small (0.24%). This indicated that the effectiveness of training was only confirmed when Japanese listeners were tested to identify [ts] vs. [tʂ] contrast. This might result from the ceiling effect of improvement, since the identification accuracy percentages for [t] vs. [t^h] was initially as high as 92.89% in the pretest in comparison to 68.13% for [ts] vs. [tʂ], which possibly limited the potential of improvement in posttests, Gen1 and Gen2 that followed.

Lastly, we want to explore whether Japanese listeners' individual language experience (in terms of length of language instruction) influenced their performance improvement across four identification tests. In addition, we also want to investigate whether the improvement between pretest and three other tests had any bearing with each other. Table 3.4 below showed all the data that were submitted to Pearson correlation tests in SPSS. Based on ANOVA results above, it was found that Japanese listeners in control group did not improve their identification accuracy across tests, so only the data for training group was included. In addition, Japanese listeners seemed to perform fairly differently in tests for two consonant contrasts, thus the improvement in terms of percentage were tested separately for each contrast.

Table 3.4. Language experiences and improvements (%) across tests for training group

Group	Subject Number	Chinese Instructions (months)	Pretest (%)for [t] vs. [t^h]	Pretest (%)for [ts] vs. [tʂ]	Improvement(%)for [t] vs. [t^h] contrast			Improvement(%)for [ts] vs. [tʂ] contrast		
					post-pre	gen1-pre	gen2-pre	post-pre	gen1-pre	gen2-pre
TR	102	3	100	50	-6.25	-12.5	0	50	0	0
TR	103	2.5	100	65.63	0	0	-3.12	25	6.25	0
TR	104	3	100	65.63	0	0	0	25	9.37	9.37
TR	105	12	43.75	81.25	53.13	21.88	40.63	15.63	3.12	3.12
TR	106	3	96.88	75	3.12	0	0	25	-12.5	-9.37

For [t] vs. [t^h] contrast pair, the correlation analysis showed that all improvements (posttest-pretest [r = .988, p=.002], Gen1-pretest [r=.891, p=.042] Gen2-pretest [r= 1, p=.000]) were significantly correlated with language experience, indicating trainees who had longer length of Chinese instruction tended to improve more greatly after the training than those who had less language experience. In addition, a significant correlation between identification accuracy improvement from pretest to posttest with two other improvements (Gen1-pretest [r=.948, p=.014], Gen2-pretest [r=.986, p=.002]) was also found. This indicated that Japanese listeners who improved greatly at posttest also tended to improve greatly at two generalization tests. Interestingly, the identification accuracy in pretest correlated negatively with the all improvements: posttest-pretest [r= -.994, p=.001] Gen1-pretest [r= -.907, p=.034], and Gen2-pretest [r= -.997, p=.000] improvement. This suggested that Japanese listeners who performed poorly in pretest tended to improve to a greater extent in later tests after training than those who performed better initially. This finding is easy to comprehend since lower pretest accuracy implied larger potential of improvement in later test sessions.

In contrast, the correlation analysis for [ts] vs. [tʂ] contrast showed that none of the improvements (posttest-pretest, Gen1-pretest and Gen2-pretest was significantly correlated with language experience. In addition, there was a significant correlation between identification accuracy improvement from pretest to Gen1 with Gen2-pretest [r=.922, p=.026] improvement, but not posttest-pretest. This indicated that Japanese listeners who improved greatly at Gen1 also tended to improve greatly at Gen2. Since the mean value of posttest-pretest improvement was much higher (28.13%) than that of Gen1-pretest(1.25%) and Gen2-pretest (0.62%), this correlation indicated that the training effect was generalized poorly, if any, to both Gen1 with novel stimuli produced by familiar talker and Gen2 with novel stimuli produced by new talker. Similar to the results of correlation analysis for [t] vs. [t^h] contrast, the identification accuracy in pretest correlated negatively with only one improvement---posttest-pretest [r= -.930, p=.022]. The interpretation is also likewise: Japanese listeners who performed poorly in pretest tended to improve to a greater extent in later tests after training.

3.3. General Discussion

In this training study, we investigated whether HVPT paradigm was effective in helping Japanese L2 learners to identify two specific Chinese sound contrasts: stop contrast [t] vs. [t^h] and affricate contrast [ts] vs. [tʂ]. The stimuli variability that is critical in HVPT paradigm was obtained by using tokens produced by multiple talkers in different syllable environments. In general, the HVPT approach was proved to be effective in shaping Japanese L2 learners' perception of Mandarin sound pairs. This training effect was not obtained by listeners' shallow memorization of token-specific phonetic cues that they were exposed during training sessions, since significant accuracy

improvement was also observed in two generalization tests, when listeners were tested on new tokens produced by either familiar or unfamiliar talker. HVPT training paradigm succeeded in drawing Japanese listeners' attention to acoustic/phonetic cues that are essential to make abstract prototypic generalization of phonetic properties of speech sounds. This finding was consistent with previous training studies using HVPT paradigm to improve listeners' perception of nonnative sounds.

The second objective of this training study was to investigate if the training effects were confirmed, whether the perception of sound pairs with different predicted discrimination difficulties were shaped differently by phonetic laboratory training. Accordingly, we specifically selected two consonant contrasts that were labeled as "hard" and "easy" in terms of discriminability respectively based on predictions laid out by PAM framework and perceived phonetic distances measured in chapter 2. However, the findings in this training study was not consistent with the prediction by PAM and was also different from the findings of the discrimination experiment in chapter 2. As discussed in chapter 2, sound contrast [t] vs. [t^h] was predicted to be "hard" to discriminate by PAM since they were both categorized into Chinese unaspirated stop [t] as "fair" exemplar and the results from discrimination experiment also confirmed this prediction. In contrast, for sound contrast [ts] vs. [tʂ], since we did not include this pair in discrimination experiment, we only have the prediction from PAM based on results on categorical mapping study. The discrimination between this sound pair was predicted to be "easy" since [$tʂ^h$] was "uncategorizable" to one single Japanese category while [tʂ] was assimilated to Japanese affricate [ts] as "good" exemplar. However, in this training study, the identification accuracy for [t] vs. [t^h] (93.56%) was significantly higher than [$tʂ^h$] vs.

[ts] (80.96%) contrast, indicating that identifying sound contrast [t] vs. [t^h] was much easier than contrast [$ʈʂ^h$] vs. [ts].

It is noteworthy that all the tests administrated in this training study were two forced-choice identification tasks in which Japanese listeners were asked to identify the sound they heard in each trial, while the tests performed in discrimination experiment in chapter 2 were two force-choice AX discrimination tasks in which listeners were asked to indicate whether the sound pair they heard were the same or different. Forced-choice identification task differs from forced-choice AX discrimination task in the way that it encourages the development of phonetic memory rather than relying on sensory memory, thus further promotes more abstract category formation instead of enhancing fine within-category acoustic differences (Jamieson and Morosan, 1986). The discrepancy in Japanese listeners' performance on distinguishing [t] vs. [t^h] contrast could possibly result from differences in focuses of identification and discrimination tasks. On the other hand, [t] vs. [t^h] contrast did not turn out to be as "hard" to discriminate as predicted by PAM in this training study either. With identification accuracy as high 92.89% at the very onset of the experiment, it is possible that the perceived phonetic distances between [t] vs. [t^h] contrast changed due to listeners' Chinese instruction. In other words, [t] vs. [t^h] were perceived to be far away from each other in terms of phonetic distances since it is likely that the category establishment for Chinese aspirated and unaspirated stops may have already come to completion or near completion at the point of testing. After all, PAM's predictions were made based on phonetic distances perceived by Japanese listeners without any Chinese experience, while the identification tests in this study were tested on Japanese L2 learners with Chinese experience varying from 2.5 to 34 months. As

indicated in the analysis of the relationship between language experience and identification accuracy above, all the improvements (posttest-pretest, Gen1-pretest and Gen2-pretest) in identifying [t] vs. [t^h] contrast was significantly correlated with language experience. This is indicative of significant influence of language experience on the discriminability between these two sounds. Different from stop contrast [t] vs. [t^h], affricate [$tʂ^h$] vs. [ts] was predicted to be "easy" to distinguish yet found to be much harder in identification tests. This discrepancy between PAM's prediction of discriminability between [$tʂ^h$] vs. [ts] and the result in identification tests in this training study reminded us of the discussion in chapter 2 about the fricative contrast [x] vs. [f]. It again reinforced the finding that PAM's predictions on UC assimilation type needs further modification when there is a overlap in assimilation category between "uncategorizable" sound and "categorizable" one. The discrimination difficulty for UC type did not turn out to be "easy" in this training study, because this overlap in assimilation categories might actually narrowed the perceived phonetic distances between these two sounds and further brought additional difficulty to discrimination. The discrepancy between PAM's predictions and the findings in identification tests in this training study made it hard to discuss the training effects with regard to different discriminability levels.

The objectives of this training study also included investigation of how the learning process of two phonetic features specific to Mandarin Chinese---aspiration and retroflexion differ from each other. The results showed that the training effects for aspiration and retroflexion differ greatly. The training was actually found to be effective only for [$tʂ^h$] vs. [ts] contrast. As indicated above, the absence of training effects for [t]

vs. [t^h] is probably due to ceiling effect since at the pretest Japanese listeners was already able to identify these two sounds correctly 92.89% of the time. Alternatively, the performance could also be indicative of the possibility that phonological categories for each Chinese stops [t] vs. [t^h] have been established or at least partially established after a few months of language instruction. This leaded us to further speculation that aspiration is a relatively easy acoustic feature for Japanese learners to acquire compared to retroflexion. As discussed in Chapter 2, although Japanese stops are distinguished by voicing while Chinese stops are distinguished by aspiration, these two features share the same acoustic correlate---VOT. In order to successfully discriminate between Chinese aspirated and unaspirated stops, Japanese listeners need to learn to shift VOT cues established for Japanese categories to Chinese-specific ranges. Adjusting the acoustic cues that is already utilized in their native language is probably much easier than trying to experiment with all possible acoustic factors to create the perceptual representation of a phonetic feature that they had never been exposed to. Alternatively, it is also possible that aspiration in Mandarin Chinese are perceptually more salient due to unique articulatory information involved in articulation. In a study evaluating the aspiration of Chinese labial stops [p^h] and [t^h] produced by Japanese L2 learners, it was found that in addition to VOT, Japanese L2 learners' production with more breathing power received higher ratings by Chinese native speakers (Hoshino & Yasuda, 2006).

Eckman (1997, 2004) have proposed Markedness Differential Hypothesis (MDH) which grounded its hypothesis on L2 learning difficulties based on a systematic comparison between L1 and L2 in the context of universal grammar. This hypothesis was widely applied and tested in various fields of linguistic studies such as syntax, phonology

etc (Chan, 2007; Jin, 2008). The difference between L1 and L2 sounds were further distinguished based on its markedness in universal grammar. For L2 sounds which are different from L1 sounds and more marked, the learning difficulty is predicted to be harder, probably due to further attention required when trying to attend to novel acoustic properties. According to Eckman's definition, both aspiration and retroflexion are both "marked" features since there exist languages in the world which have neither of these features. However, for marked features, there seems to exist a more fine-grained ranking of markedness. In this case, retroflexion was found to be more marked than aspiration and thus more difficult to learn.

In contrast to stop contrast [t] vs. [t^h], the same amount of Chinese instruction apparently failed to place Japanese listeners at the same stage of category establishment for affricate contrast [$ʈʂ^h$] vs. [ts]. The lack of correlation between language experience and improvements from pretest to other three tests (posttest, Gen1 and Gen2) also reinforced our speculation. Retroflexion, as a novel and more difficult phonological feature probably needs longer time of instruction which can provide systematic input that can expose Japanese learners to substantial acoustic cues essential for characterization and further category establishment, which is exactly what was provided in our training session. In a perception study examining Mandarin and Taiwanese listeners' perception of Chinese alveolar-retroflex contrasts [s, ts, ts^h] vs. [ʂ, ʈʂ, $ʈʂ^h$] produced in different vowel contexts ([a] and [u]) (Chang, Shih and Allen, 2013), it was found that retroflex consonants had displayed more complex and varied acoustic properties than alveolar consonants. Chinese native speakers had showed much more tolerance to sub-phonemic variation of the retroflex category than alveolar category, which means that all variants of

retroflexes were perceived by Chinese Mandarin speakers as equally good. This acoustic difference in alveolar-retroflex contrasts was further explained by the authors from an articulatory perspective. The articulation of alveolar is very restricted since its articulation only involved raising the tongue tip. In contrast, more complex tongue configurations such as blade or lip rounding were required when producing retroflex, thus result in a high degree of variance in production. This can account for the difficulty Japanese listeners were facing in distinguishing [ts] vs. [tʂ] in this study. In speech perception, in order to form a new category, abstract prototypic information has to be extracted from acoustic signals full of detailed token-specific information such as talker's voice and different surrounding phonetic environments etc. This abstraction process enables the categorization to be invariant to different acoustic factors and then can be stored in long-term memory (Pisoni, 1992). In the case of current study, in order to acquire phonetic feature of retroflexion, Japanese listeners are faced with a difficult task of perceptually filtering through all the detailed acoustic properties of the speech signals they were exposed and grasp the essential acoustic cues for retroflexion categorization. This account, on the other hand, also explained why HVPT paradigm in this training study was successful. In training session, high stimulus variability served as the essential sources that provided substantial variations of speech signals from which prototypes can be extracted to characterize the complex phonetic profile of Chinese affricate retroflex .

Another explanation for the difficulty faced by Japanese listeners when trying to distinguish [tʂ^{h}] vs. [ts] is possibly due to the phenomenon of retroflex and non-retroflex merger in Mandarin Chinese. Mandarin Chinese spoken by Chinese native speakers from some regions of China, usually southern China (e.g. Shanghai) are merging retroflex and

non-retroflex in their speech thus making [tʂʰ] vs. [ts] undistinguished (Zhu, 2012). If Japanese listeners are exposed to Mandarin Chinese speech that had this merger during their experience of learning Chinese, it is highly possible that the speech input with affricates lacking the distinction of retroflexion would make it much harder for L2 learners to establish category for this phonetic feature.

Chapter IV

ACOUSTIC ANALYSIS OF PERCEIVED ACCENTEDNESS IN JAPANESE L2 LEARNERS' PRODUCTION OF MANDARIN CHINESE

4.1 Introduction

The previous chapters (Chapter 2 and Chapter 3) investigated how Japanese L2 learners perceive Mandarin Chinese sounds and whether their perception can be improved by laboratory training. This chapter presents the examination of Japanese L2 learners' learning of Mandarin Chinese from the perspective of speech production. The most straightforward criterion to evaluate L2 speech is foreign accent, which has been bothering L2 learners from the very onset of L2 acquisition. The phenomenon of foreign accent in second language (L2) learners' production has long attracted researchers' interests (Piske, MacKay &Flege 2001 for review). Most of the research efforts in L2 studies have been devoted into examining L2 learner-related factors contributing to the perceived foreign accentedness. Among a broad range of factors examined, many studies converge in demonstrating that the onset age of learning (AOL) exerts a crucial influence on the development of a perceived foreign accent (Long, 1990; Oyama, 1976; Scovel, 1988; Tahta, Wood, & Loewenthal, 1981), whereas studies are inconclusive about the influence of other factors, such as length of residency (LOR) in the community where the language is spoken gender, formal instruction, motivation and so forth (Elliott, 1995; Moyer, 1999; Thompson, 1991). This study, however, will focus on examining the production of L2 learners in terms of segmental and suprasegmental features and

investigate the relationship between acoustic characteristics of the L2 production and accentedness ratings provided by native listeners.

This Chapter aims to contribute to the research on foreign accent in SLA by examining one of less-studied language pairs---Mandarin-Japanese, specifically for Japanese speaking L2 learners' production of Mandarin Chinese. The primary objective is to contribute to the understanding of the acoustic correlates of a foreign accent, the theoretical implications in terms of SLM and PAM and how the findings can inform non-native speakers in terms of reduction of the degree of perceived foreign accentedness of their speech.

The research questions addressed in this chapter are as follows: What are the relative weightings of segmental or suprasegmental variables on the perception of foreign accent? In each category, which specific variables are the most influential? The theoretical implication of the results in terms of theoretical second language learning model (PAM and SLM) will also be addressed in the discussion section.

4.2. Acoustic Sources of Foreign Accent

In contrast to speaker characteristics, acoustic properties of non-native speech that give rise to the perception of a foreign accent are far from well-explored. A few published studies that have examined acoustic correlates of foreign accents reveal that there are certain salient acoustic features that seem to affect perception of a foreign accent more than others (Wayland, 1997; Munro, 1995; Trofimovich& Baker, 2006).

In a small scale study, Wayland (1997) examined potential acoustic source of foreign accent on American learners' L2 production of Thai. She examined two vowels (monophthong [a:] and diphthong [a:u]) and one consonant ([k^h])produced by 3 Thai

native speakers and 6 male native American learners differing in years of learning experiences measured in LOR (6 weeks to 3.5 years). The stimuli were elicited by having participants read a wordlist of two strings ([k^ha:u] and [na:]) in five tones variations (low, mid, high, falling and rising).Acoustic measurements revealed that out of all the variables examined---vowel formants (F1, F2, F2-F1), fundamental frequency (F0) peak (the highest F0 value on the F0 contour), F0 valley(the lowest F0 value),F0 range(the difference between F0 peak and F0 valley), stop VOT and vowel duration, American learners differed from native speakers in terms of spectral parameters, such as F1, F2 value and F0 valley rather than temporal parameters, such as VOT and vowel duration. Productions from both native speakers and American learners group were further presented to 3 male Thai native speakers for evaluation of degree of accentedness based on 5-point scale (1 indicated strong foreign accent and 4 indicated native Thai). It was found that not all acoustic differences were directly translated to perception of foreign accent. Multiple regression analysis on the correlation relationship between acoustics and accented ratings revealed that only F0 valley and F2 value of [u] in [a:u] were consistent in predicting foreign accent for all five tones, while other predictors varied from tone to tone. This study also found that language experiences did not help in eliminating degree of foreign accentedness. Although only a very limited Thai phonological inventory was examined in this study, the findings still confirmed that foreign accent ratings were influenced by both segmental (e.g. F2 values) and suprasegmental factors (e.g. F0 valley).

Whereas Wayland (1997) examined both segmental and suprasegmental features, Munro (1995) attempted to focus on suprasegmental features in Chinese learners' L2

English. Speech sample,consisting of read sentences, were obtained from 10 native English speakers and 10 native Mandarin speakers. In order to examine the exclusive role played by suprasegmental factors on foreign accent, these stimuli were then low-pass filtered, 225 Hz for male voices and 300 Hz for female voices, to eliminate segmental information while retaining most of suprasegmental information (e.g., F0, word duration and rhythmic properties). Orignal and filtered stimuli were evaluated on a 4-point scale, with a higher point indicating more native-like production.Results revealed that native English speakers received statistically significant higher rating scores (ranging from 1.4-2.5) than Mandarin speakers (ranging from 2.4-3.5) for filtered stimuli. The lack of overlap in these ranges of accentedness rating scores between English speakers and Mandarin learners were taken to mean that suprasegmental information alone was sufficient to distinguish foreign accented speech from native speech.While this study does suggest an important role of suprasegmental features on perceived foreign accent, relative importance of suprasegmental and segmental features is still not clear. Also since the accent rating was not related to acoustic measurements of the speech samples, the acoustic sources of the perceived accentedness is not clear. Nevertheless, this study served to present an effective technique to separatethe suprasegmental from segmental information in speech materials and suggest an important role played by suprasegmentals in the perception of foreign accent.

The method of low-pass filtering was employed also by a more recent study by (Trofimovich & Baker, 2006) who noted the scarcity of research on L2 learning of suprasegmenals and analyzed the role of several suprasegmental features on perceived L2 accentedness. Production of 6 predetermined sentences were elicited from three groups of

Korean learners of English differing in LOR (3 months, 3 years, and 10 years respectively) and one native English speaker group. Another 10 English native speakers evaluated low-pass filtered productions from the speakers for the degree of foreign accent on a 9-point scale. Making an improvement over Munro (1995), this study then related the accent rating to acoustic measures of five suprasegmental features: stress timing (measured by stressed and unstressed syllable-duration ratios), tonal peak alignment(measured by the duration between the onset of the vowel in the stressed syllable and the highest value of fundamental frequency in an intonation phrase), speech rate (measured by number of miliseconds per syllable one utterance), pause frequency and duration (the average number of pauses and pause durations across all sentence stimuli).The results from acoustic analysis alone confirmed the effects language experience---LOR (Length of Residence) on stress timing as well as the effect of AOA (Age of Acquisition) on speech rate, pause duration and pause frequency. More importantly, a regression analysis relating the suprasegmental measures and accent rating showed that pause duration and speech rate have more influence on perceived accent than the other three variables, pause frequency, stress timing and peak alignment. Given this the researchers reported that fluency characteristics (e.g., pause duration and speech rate) may affect perceived foreign accent than melodic characteristics (e.g., stress timing and peak alignment).

These studies reviewed above and others (Anderson-Hsieh, Johnson, & Koehler, 1992) indeed suggest an important role that suprasegmental features play in the perception of foreign accent in L2 speech. Furthermore, these studies in the process developed methodology allowing us to investigate the role of suprasegmentals on

perceived accent (i.e., using filtered speech sample) and relating different acoustic features to accent ratings (i.e., using regression analysis). However what is still lacking is an examination of the relative importance of suprasegmentals vis-a-vis that of segmentals. While it is impossible to separate those major features of speech from each other while maintaining the integrity of speech materials, a comparison of suprasegmentals and segmentals is an important task with considerable pedagogical implications. Since in L2 or foreign language classroom, instructions on pronunciation usually focus on segmental features of the target language, identifying the relative weights carried by both segmental and suprasegmental features in terms of foreign accent perception will inform language teachers of effective pedagogies The current study attempts to approximate a comparison of these two important domains of speech by using both filtered and unfiltered speech samples in our investigation.

Another important consideration for research on this topic is whether the acoustic correlates for accent perception are universal across different languages or whether they are language-specific. It is not difficult to imagine that what comprises a sense of accent may be different depending on the target L2 language, while it may also depend on the paring of target L2 and learners L1. At the same time, it is also conceivable that, given the common cognitive and auditory faculty, we all detect foreign accent in some similar ways cross-linguistically. We cannot begin to address this question until we have a database of studies examining multiple languages and language pairings between the target language and native language. The current study also attempts to contribute to knowledge base by examining less-commonly examined pair – L2 learning of Japanese by Chinese (Mandarin-speaking) learners.

4.3. Production Study

4.3.1 Methods

4.3.1.1 Participants

Twenty-three Japanese L2 learners of Chinese (10 female and 13 male) and 10 native-Chinese speakers (7 female and 3 male) provided speech samples. All Japanese L2 learners were native speakers of Japanese, speaking the language since birth. The mean age of the learners was 23.6 (range: 19-39). They were enrolled in either elementary (4), intermediate (10), and advanced (9) level Chinese language course at the Beijing Language and Culture University at the time of testing. Learners were recruited from these levels so that the speech samples included a range of Chinese language proficiency. All of the Japanese L2 learners learned a foreign language before age 14, except for two who learned English at age 4 and 12. More information regarding the language experience of these learners is provided in Table 4.1.

All 10 native Chinese speakers who provided speech samples were from mainland China. Their mean age was 28.5 (range: 26-31) and all were attending the University of Oregon at the time of testing. Their average length of stay in the US was 4 years (range: 2 years – 5.5 years). All reported daily use of Mandarin Chinese.

Table 4.1. Language background of L2 participants

	Japanese L2 Learners (n = 23)
Age (years)	23.6 (19-39)
AOA (years)	21.2 (16-37)
Japanese instruction (years)	4.6 (2-8)
Length of stay in China (months)	13.6

Table 4.1. Language background of L2 participants (continued)

Lengths of stay in China	**Number of L2 learners**
Less than 1 month	6
1-5 months	2
6-11 months	5
more than 1 yr	8
more than 3 yr	2

4.3.1.2. Materials

The test materials used to elicit speech samples were 6 Chinese sentences (Appendix C). The vocabulary and sentence structure were taken from a beginning level Chinese language course book *Hanyu Jiaocheng* (Jizhou Yang, 1999) and adapted for the length and comprehensibility appropriate for participants as well as for the range of segments included in them.

Learner and native-speaker productions of the test sentences were collected using delayed repetition task, an elicitation technique used in L2 research as a method that allows relatively natural elicited production while maintaining the control of the speech materials (Flege et al, 1995; Trofimovich and Baker 2006). Two native Chinese speakers, a female (the author) and a male, recorded 6 prompts for the task (Appendix C), each comprising a question-response-question sequence as shows in (1). In each sequence, the response is one of the six test sentences.

(1) Question (male): *Na shi shen me za zhi?* "What kind of magazine is that?"
Response (female): *Na shi ying wen za zhi.* "That is English magazine."

Question (male): *Na shi shen me za zhi?* "What kind of magazine is that?"

As the example shows, the first speaker asked a question, followed by a response by the second speaker. Then, the first speaker repeated the question. The repetition of the response was not included in the prompt so that the participants would be prompted to

produce the response. This design allows elicitation of target utterances while avoiding immediate repetition of the model (Piske et al., 2001). The six prompts for the task were recorded in a quiet room using a flash digital recorder (Tascam DR-100mkII - Portable 2-Channel Linear PCM Recorder) and a microphone (Audio-Technica AT8537) at a sampling rate of 44 kHz and 16-bit quantization. These two native speakers only recorded the task prompts and did not participate in the subsequent study.

4.3.1.3 Production Task

The six recorded prompts were presented to participants auditorily and participants were recorded producing the test sentences as the response to the question. Each participant was first provided with two practice prompts followed by the six test prompts in a random order, with each sequence presented three times consecutively. The recordings were conducted individually in a quiet room with the same setting used for recording the prompts. The experimenter (the author) was present in the room during the recording to present the prompts using Apple iTunes software to generate randomized playlist of stimuli.

4.3.1.4. Segmental Variables and Measurements

The current study focuses on stop consonants and vowels in the segmental domain based on cross-linguistic comparisons as discussed in Introduction. All the measures reported in this chapter were taken from either the third or fourth repetition the target sentences. When the third repetition included disfluency (e.g., wrong word, false start), the forth repetition was selected.

We examined the duration of closure and VOT in stops (i.e., [t, t^h, p, p^h, k, and k^h]) and the frequency of the first formant (F1) and the second formant (F2) of vowels)

were measured in vowels (and their allophones) [i, ɨ], [y], [æ, a, ɑ], [u], [e, ə, ɤ, o], and semivowels [w, j, ɥ] in the mid-point of each vowel by using waveform and spectrographic displays in Praat 5.2.18 (P Boersma & Weenink, 2005). Stop closure was measured from the offset of the preceding vowel to the release of the stop burst, while VOT was measured from the release of the stop burst to the onset of the following vowel with reference to visible F2 energy at the spectrogram (Idemaru & Guion-Anderson, 2010) Stops were sometimes preceded by a phrase boundary where a pause could occur. When a stop is preceded by a pause, it is not possible to identify the beginning of the stop which begins with a closure. Thus, stop closures exceeding 100 milliseconds were classified as a pause (Trofimovich and Baker, 2006), and were excluded from the analysis of stop duration. Frequency of F1 and F2 was measured for each vowel (and their allophones) [i, ɨ], [y], [æ, a, ɑ], [u], [e, ə, ɤ, o], and semivowels [w, j, ɥ]. The vowels formants were then normalized for individual differences using the Lobanov method (Nearey, 1978; Thomas & Kendall, 2007).

4.3.1.5. Suprasegmental Variables---Rhythm Measures

Also based on cross-linguistic comparisons as discussed in Introduction, current study examined several suprasegmental variables, including global rhythm, syllable timing, accent and intonation, and fluency.

The global rhythm measures used in this study were as follows: ΔV and ΔC, which index the standard deviation of the duration of the vowel and consonant intervals in each utterance; V%, the percentage of the total duration of vowels in each utterance (Ramus, Nespor, & Mehler, 1999); VarcoΔC and VarcoΔV[1], similar to ΔV and ΔC but

[1]VarcoΔC is 100 times ΔC divided by the mean vowel duration. VarcoΔV is 100 times ΔV divided by mean vowel duration in one sentence.

corrected for speech rate (Dellwo, 2006; White & Mattys, 2007); nPVI[2] (normalized Pairwise Variability Index), the deviations of durations of vowels in adjacent syllables excluding the influence of speech rate (Grabe & Low, 2002). According to Ramus et al (1999), ΔC and %V most successfully classify different linguistic rhythm patterns: stress-timed, syllable-timed and mora-timed. Compared to stress-timed English and Dutch, mora-timed Japanese has smaller ΔC and larger %V due to its limited syllable types (Ramus et al, 1999). Since nPVI measures the variations of durations of the vowels in adjacent syllables, syllable-timed languages tend to have smaller nPVI value while stress-timed language has greater value (Thomas Erik, 2011). Chinese is considered syllable-timed (Mok, 2009) while Japanese is classified as mora-timed (Vance, 2008). Japanese learners of Chinese may show non-native patterns in their linguistic rhythm.

4.3.1.6. **Suprasegmental Variables---Intonation measures (C_ToBi)**

Tone and intonation: We needed a framework to characterize tonal patterns of the speech samples. C_ToBi (Chinese Tone and Break Index, Li, 2002) was adapted for that purpose and used to describe lexical, phrasal and sentence-level tonal patterns. In ToBi systems, tiers are used to analyze different aspects of prosody. This version of C_ToBi system (Li, 2002) identifies (1) Pinyin tier (1, 2, 3, 4 and 0 labeled for Chinese four canonical lexical tones and one neutral tone), (2) Tone and intonation tier, which codes lexical tone labels (H-L, L-H, H-H, L-L and H/L for Chinese four canonical tones and one neutral tone), four boundary tones (%H, %L, H% and L%) indicating the start and end point of a prosodic unit, upstep/wide upstep, downstep/wide downstep labels (^,

[2]Computed as the sum of the absolute values of the differences of durations of vowels in adjacent syllables, divided by the mean duration of each pair of two adjacent vowels minus one.

nPVI $=100 \times \sum_{k=1}^{m-1}\left(\frac{(dk-dk+1)}{\frac{dk+dk+1}{2}}/(m-1)\right)$ where m is the total number of vowels in the utterance and d^k is the duration of the k^{th} vowel.

^^, !, !!) etc), and four pitch register change labels (Re^() Re^^(), Re!() Re!!() indicating the direction of pitch register change (e.g. upstep, wide upstap, downstep, wide downstp) and its starting and ending point, (3) break indices (0, 1, 2, 3, 4, 1p, 2p and 3p) to code the length of the perceived pauses between two syllables (i.e., words), (4) stress index tier (1, 2, 3 and 4) to indicate hierarchical stresses corresponding to prosodic units, (5) sentence function tier to indicate interrogative, imperative, declarative and exclamation sentences, (6) accent tier to indicate regional accent, (7) turning taking tier to give the start and end point of each turn and (8) miscellaneous tier to code paralinguistic and non-linguistic phenomena. As our stimulus sentences were short and simple, only a subset of these categories was necessary to characterize the tonal patterns of the speech materials. The following prosodic aspects were selected and coded to analyze the speech materials: (1) lexical tone (H-H, L-H, L-L, H-L) and neutral tone (H/L), (2) 5 break indices (0, 1, 2, 3 and 4) to characterize pauses, and (3) intonational boundary tones (%H, %L, H%, and L%).

Four canonical lexical tones of Mandarin (#1 above) include a high level tone (H-H), a high rising tone (L-H), a low falling-rising tone (L-L) and a falling tone (H-L). The pitch height of a neutral tone was determined by the tone of the preceding syllable (H/L). Break indices are used to code the length of perceived pauses between two syllables. Break index 0 indicates the minimum break between syllables, usually occurring within prosodic words, for example, between *za* and *zhi* in *zazhi* ("Magazine"). Pauses within prosodic words are not expected in fluent speech. The prosodic word *zashi* ("Magazine") may combine with another, *yingwen ("English"),* to create a phrase *yingwen zashi* ("English magazine"). It is not typical, but there may occur a pause between two

prosodic words within a phrase (thus, between *yingwen* and *zashi*, for example). Such a pause in my speech sample was coded with Break index 1. At the next level, there can be a phrasal boundary between two prosodic phrases. A prosodic phrase *yingwen zashi* ("English magazine") can be combined with another, *na shi* ("That is"), to create a longer phrase *na shi yingwen zashi* ("That is English Magazine"). When there was a minor pause between the two prosodic phrases (between *na shi* and *yingwen zashi*), the pause was coded with Break index 2. If the pause between these two prosodic phrases was a major (i.e., longer) pause, it was coded as Break index 3. At the next higher lever, a grouping of one or more prosodic phrases creates a prosodic group. The pause between the prosodic group boundary was coded as Break index 4. For instance, in sentence 5, *Bu, ta men bu shi ri ben liu xue sheng, ta men dou shi zhong guo xue sheng.* ("No, they are not Japanese international students, they are Chinese students") The pauses between *ta men bu shi ri ben liu xue sheng* (No, they are not Japanese international students) and *ta men dou shi zhong guo xue sheng* (they are Chinese students) indicated the separation of two distinct prosodic groups and thus was coded as Break index 4.

While break indices were used to code pauses, intonational boundary tones were used to code tonal patterns of prosodic groups. A prosodic group could begin with either high tone or low tone (%H or %L), and could end with a high tone or a low tone (H% or L%). Whether a prosodic group starts or ends with a %H or %L tone is determined by the lexical tone of its first and last syllable. For instance, in sentence 1, *Na shi yingwen zashi* ("That is English Magazine"), since the lexical tone of the initial syllable *na* is a falling tone H-L and final syllable is also a falling tone H-L, this sentence starts with %H boundary and ends with a %L boundary tone.

In native speakers' speech sample, a pause (silence duration longer than 100ms) was typically the indicator of a new prosodic group. Such pause was labeled with Break index 4, and the following tone was labeled as a new intonational boundary tone (%H or %L). However, in some cases, especially in Japanese L2 learners' production, some pauses occurred due to disfluency, accompanying self-correction, fillers or hesitation, and did not introduce a new prosodic group. In these cases, no new intonational boundary tone was assigned to the beginning of the following phrase.

Chinese native speakers' productions for each target sentence was first labeled with these aspects of tone and intonation, and used as the target patterns. Japanese L2 learners' production was then labeled and evaluated against the native speakers' patterns. Each label in Japanese learners' production for each sentence that matched the labeling in native speakers' production was given one score. For each speaker, the tone and intonation score was computed by averaging all scores across six sentences. While in general Chinese native speakers displayed consistent tonal patterns, there did exist some variations. For example, stimulus sentence 2 (*Bu. Ta men bu shi ri ben liu xue sheng. Ta men dou shi zhong guo xue sheng.* "No, they are not Japanese students. They are Chinese students.") showed several variations in terms of the lexical tones (some Chinese native speakers applied tone sandhi rule and used neutral tone while some did not) and different break indices (some Chinese native speakers paused longer than others between certain syllables). In this case, the labeling of learners' production that matched any variation of native speakers' labeling was given one score.

4.3.1.7. Suprasegmental Variables---Fluency Measures

It has been proposed that fluency characteristics may have an important role in

perceived accentedness (Trofimovich and Baker, 2006). Six variables characterizing production fluency, i.e., speaking rate, articulation rate, pause duration, pause frequency, the number of false start and self-correction were examined.

Speaking rate, the rate of speech including the pauses, was computed by dividing the duration of each utterance including pause, false start and self-correction durations (ms) by the number of syllables of the sentence. Similarly, articulation rate was computed by dividing the duration of each utterance excluding pause, false start and self-correction durations (ms) by the number of syllables of the sentence. Pause was defined as a silent period within an utterance that was longer than 100 ms (Trofimovich and Baker, 2006). Pause duration was computed for each speaker as an average of all pauses across 6 sentences. Pause frequency was an average number of pauses across 6 sentences for each speaker. Similarly, the number of false start and self-correction was also computed by averaging the number of false start and self-correction across 6 sentences for each speaker.

4.3.2 Results

4.3.2.1 Segmental Variables

This section presents descriptive statistics of Japanese learner group and Chinese native speaker groups' production to report the segmental measurements. Statistical tests were not conducted to compare the learner and native speaker values since a large number of comparisons is likely to cause type II error. The primary focus of this study is relating the acoustic measurements and the degree of perceived foreign accent, and this is presented in the next section. This section serves as an introduction to the measurement data.

The mean values of normalized F1 and F2 formants for vowels are reported in Figure 4.1, 4.2 and 4.3. Figure 1 presents high front vowels, Figure 4.2 presents back and low vowels, and Figure 4.3 presents mid vowels. In general, the differences between Japanese learners' and Chinese native speakers' production of Mandarin vowels seem to be larger for F2 than F1. As we can see in Figure 4.1, Japanese learners tended to produce [ɨ] with a further front tongue position (mean F2 values for L2 and NS: 0.39 vs. 0.21) and [y] (mean F2 for L2 and NS: 0.44 vs. 0.70) with a further back tongue position than Chinese native speakers.

Figure 4.2 shows that, Japanese learners tended to produce vowel [u] with a further front tongue position (mean F2 for L2 and NS: -0.82 vs. -1.08) and [a] with a further back tongue position than Chinese native speakers (mean F2 for L2 and NS: -0.39 vs. -0.14). Figure 4.3 shows that Japanese learners tended to produce mid-central vowel [ə] with a further front tongue position (mean F2 for L2 and NS: 0.23 vs. 0.02) and produce mid-back vowel [o] with a further back tongue position than Chinese native speakers (mean F2 for L2 and NS: -1.07 vs. -0.86).

The mean values of stop duration measurements for Japanese L2 learner and Chinese native speaker groups are reported in Figure 4.4 below. Interesting differences may be in the production of aspirated stops. Japanese learners' VOTs appeared to be shorter and their closures appeared to be longer compared with Chinese native speakers. The differences in the production of unaspirated stops were less obvious. It seemed that Japanese L2 learners produced slightly longer VOTs and shorter closures for unaspirated stops.

Figure 4.1. Mean vowel formant values for vowel [i, j, ɥ, y, ɨ, ɚ] (Lobanov normalized)

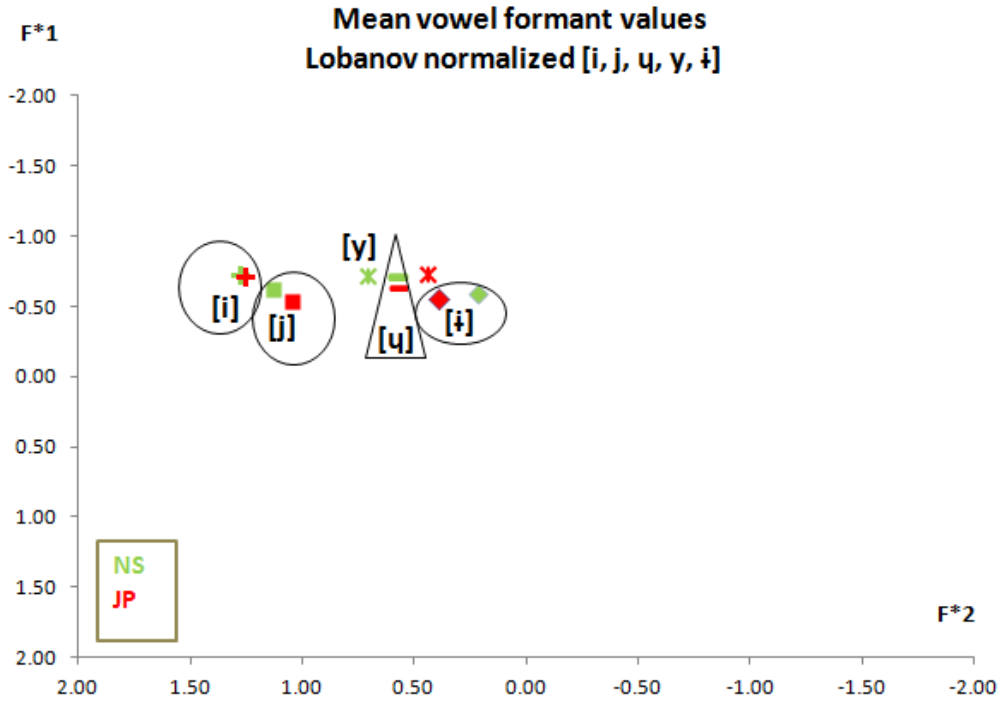


Figure 4.2. Mean vowel formant values for vowel [æ, a, ɑ, u, w] (Lobanov normalized)

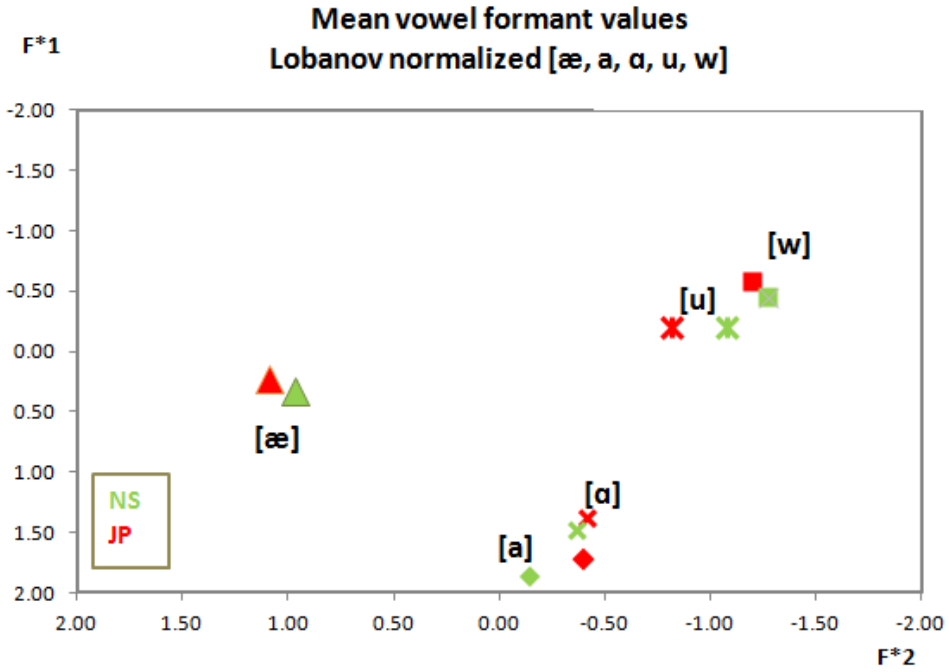


Figure 4.3. Mean vowel formant values for vowel [e, ə, ɤ, o] (Lobanov normalized)

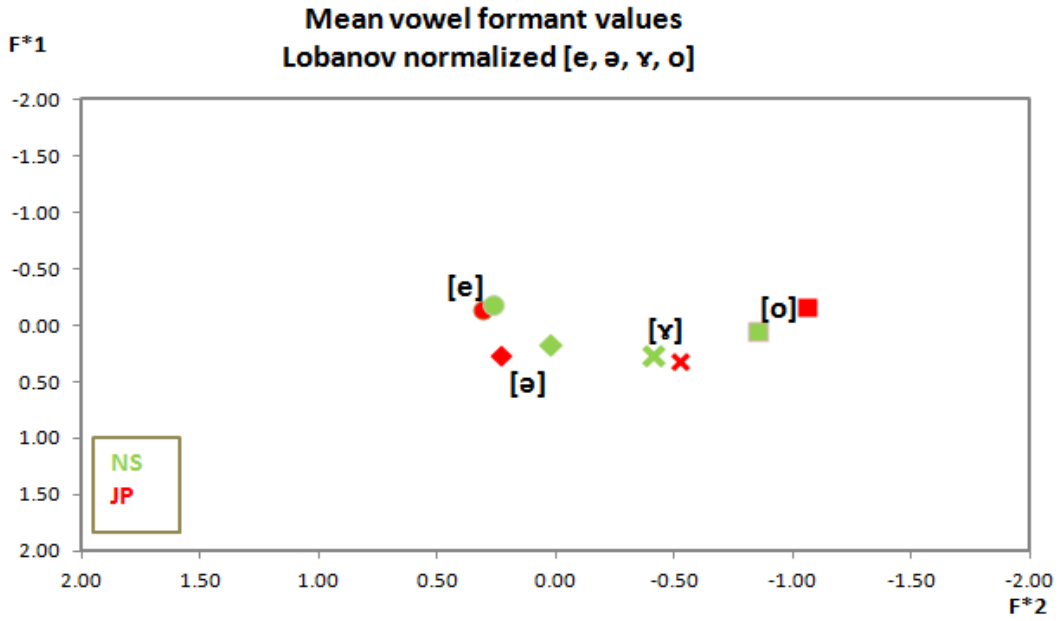
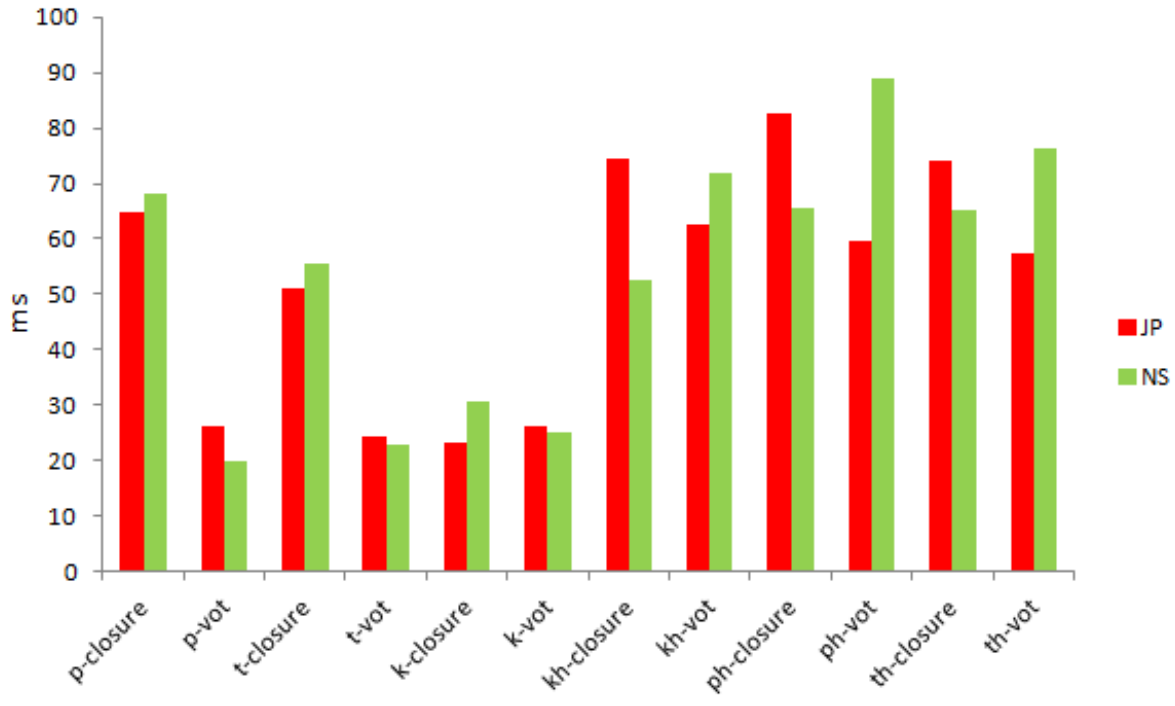


Figure 4.4. Closure and VOT durations for aspirated [p^h, t^h, k^h] and unaspirated stops [p, t, k]



4.3.2.2 Suprasegmental Variables

The mean values of ΔV, ΔC, VarcoΔV, V% , VarcoΔC, nPVI, syllableSD (ΔSyllable) and C_Tobi score for Japanese L2 learner and Chinese native speaker groups are reported in Figures 4.5-4.8. The measures of six variables characterizing production fluency: speaking rate, articulation rate, pause duration, pause frequency, the number of false start, self-correction are reported in Figure 4.9-4.12.

Figure 4.5. V% and nPVI

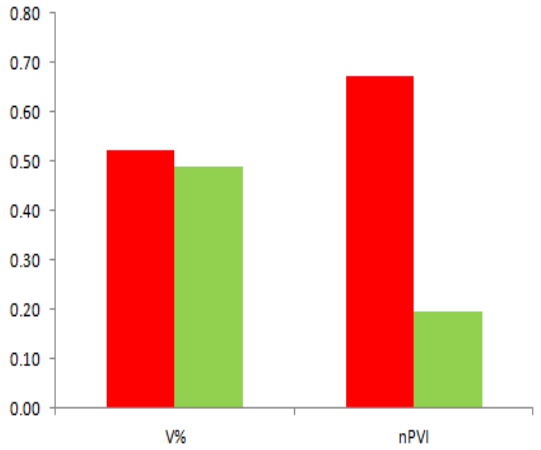


Figure 4.6. Standard Deviation of Syllable

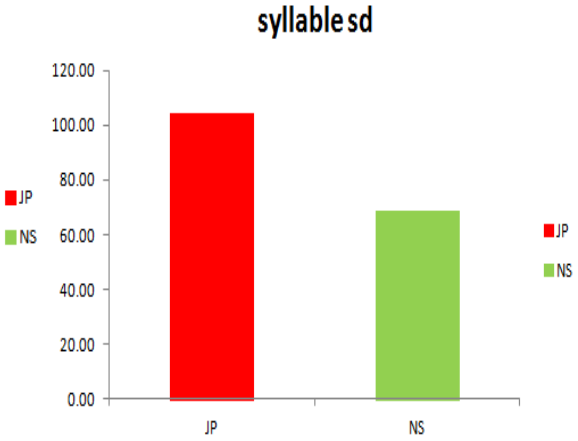


Figure 4.7. ΔV, VarcoΔV, ΔC, VarcoΔC

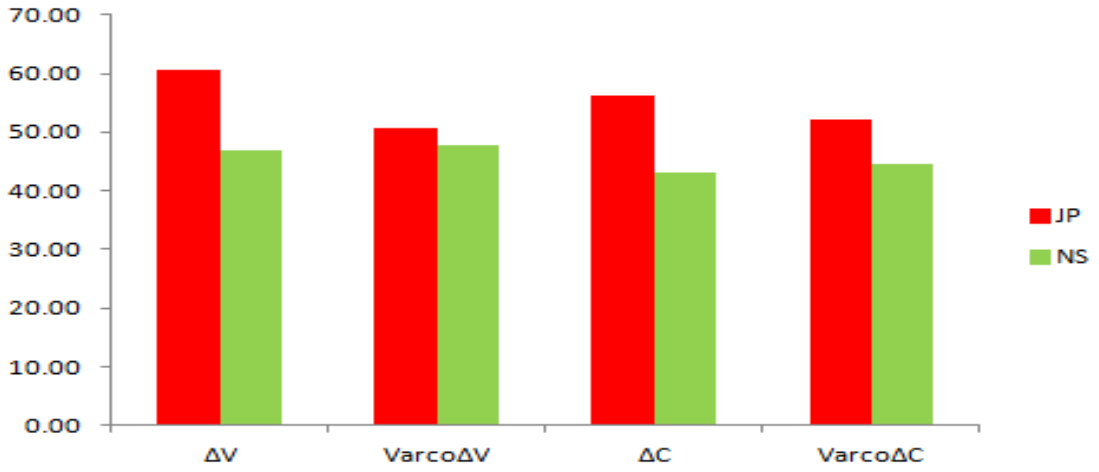


Figure 4.8. Chinese ToBi Score

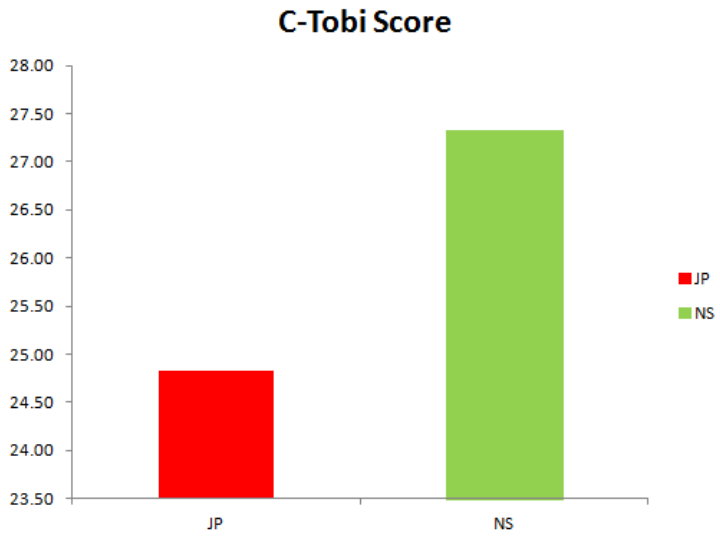


Figure 4.9. Articulation Rate and Speaking Rate

Figure 4.10. Pause Duration

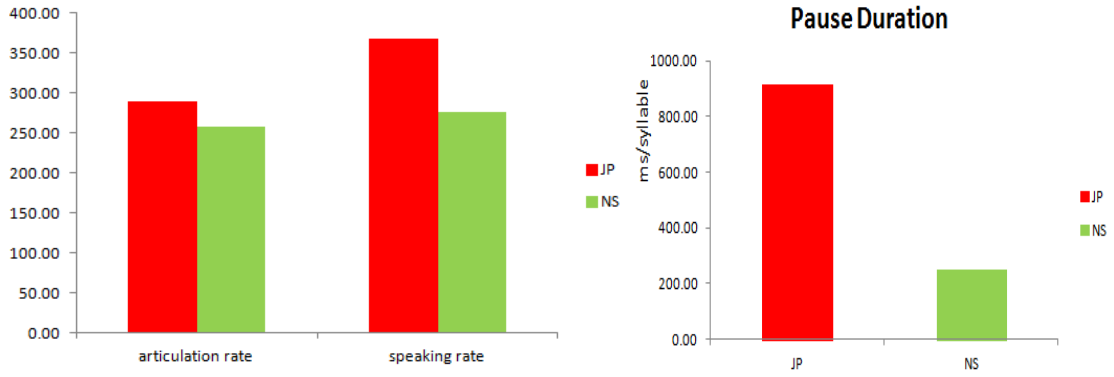
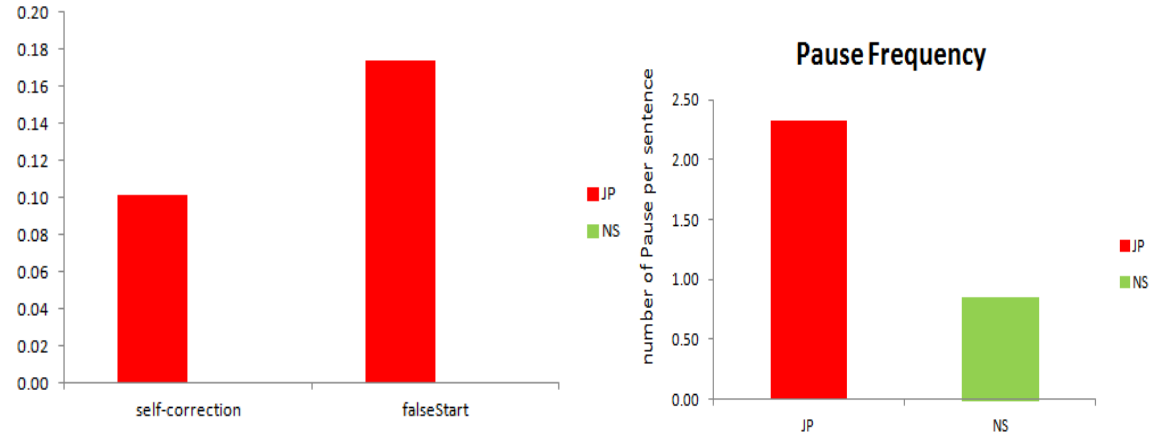


Figure 4.11. Self-correction and False Start

Figure 4.12. Pause Frequency



In Figure 4.5, the difference in V% seems to be small, but the robust difference in nPVI indicates that the duration of adjacent syllables was more consistent for native Chinese speakers and more variable for L2 speakers. Measures of durational variability in Figures 4.6 (Syllable SD) and 4.7 (ΔV, VarcoΔV, ΔC, VarcoΔC) illustrate that there are greater variations in syllable durations (Figure 4.6), vowel durations and consonant durations (Figure 4.7) of Japanese L2 learners than Chinese native speakers' production. Chinese ToBi scores reported in Figure 4.8 showed that Japanese learners spoke with tonal and intonation patterns that are different from Chinese native speakers

Fluency measures in Figure 4.9-4.12 indicated that Japanese L2 learners spoke more slowly (Figure 4.9) with longer and more frequent pauses (Figures 4.10 and 4.12) and more frequent false starts (Figure 4.11).

4.3.3 Discussion

As shown in the results above, Japanese L2 learners tended to produce higher F2 value (thus with further front tongue position) in high front vowel [ɨ], high back rounded vowel [u], and mid-central vowel [ə], but lower F2 value (thus with further back tongue position) in high front rounded vowel [y], mid-back vowel [o], front low vowel [a].

I will note here that the non-native-like production of Chinese [u] was predicted by SLM and the results of the mapping study (Chapter 2). Japanese learners were likely to have classified the Chinese [u] as equivalent to Japanese [ɯ], as the participants did in the mapping study. According to the results of the categorical mapping experiment in Chapter 2, Chinese [u] was perceived to be a good exemplar of Japanese [ɯ] since it was classified into Japanese [ɯ] for 95% of all the tokens, and received a goodness-of-fit score as high as 3.79 out of a scale of 5, yielding a fix index of 3.60. Phonetic details of

Chinese vowel [u] were not picked up by Japanese L2 learners due to this equivalence classification and eventually resulted in resemblance of Japanese [ɯ] in their speech production. For Chinese central vowel [ə], as indicated in mapping study, there is no high degree of consensus with regard to which Japanese category Japanese L2 learners would classify for Chinese central vowel [ə]. This vowel was classified into eight different Japanese vowel categories and three most frequently identified Japanese categories were [o] (30%), [e] (27%) and [ɯ] (26%) with goodness-of-fit ratings of 3.04, 2.58 and 2.30 respectively. Based on the results of categorical mapping study in Chapter 2, we can see that Chinese central vowel [ə] was perceived to be "poor" exemplar for three Japanese vowel categories [o], [e] and [ɯ]. According to the hypothesis of SLM, a new phonetic category of a L2 sound can be established if the phonetic dissimilarities between this L2 sound and its closest L1 counterpart can be discerned. Also, the production of a L2 sound eventually corresponds to properties of the phonetic category established for this sound. Accordingly, instead of simply using one preexisting Japanese vowel category as the equivalence category, Japanese L2 learners might have established a category for Chinese central vowel [ə] in their vowel space because it is not similar to any Japanese categories. However, the properties of this newly established category differed from that of Chinese native speakers' production, thus led to discrepancy in Japanese L2 learners production of this sound. Similarly, Chinese front rounded vowel [y] was also most frequently classified as a poor exemplar of two different Japanese categories [ɯ] and [i]. Establishing a new phonetic category for Chinese [y] helped L2 learners to distinguish this vowel from other Japanese categories, yet did not help Japanese L2 learners approximate their production to that of Chinese native speakers, since the phonetic

features of this new category established by Japanese L2 learners usually diverge from that of Chinese native speakers. For Chinese front low vowel [a], it was perceived as a "good" exemplar of Japanese category [a]. Similar to Chinese vowel [u], phonetic differences between Chinese [a] and Japanese vowel [a] were blocked due to the closeness of phonetic distance between these two categories. Although it needs further investigation with regard to whether Japanese [a] has lower F2 values than Chinese [a], this lack of perceptual differentiation between these two categories was probably also reflected in Japanese L2 learners' production, leading to difficulty in producing native-like Chinese [a].

As for stop production, Japanese L2 learners showed more systematic tendency in aspirated than unaspirated stops by producing shorter VOTs and longer closures for aspirated stops [p^h, t^h, k^h], but slightly longer VOTs and shorter closures for unaspirated stops [p, t, k] compared with Chinese native speakers. These learners did not seem to have good control of VOT appropriate for Chinese stops. These results confirmed the prediction we made in the Introduction section that Japanese L2 learners would have great challenges in producing native-like Chinese aspirated stops and unaspirated stops.

According to the results of categorical mapping study in Chapter 2, Chinese unaspirated stops [p, t, k] were mapped to Japanese voiceless stops [p, t, k] as “fair” exemplars, while Chinese aspirated voiceless stops [p^h, k^h] were mapped to Japanese voiceless stops [p, k] as “good” exemplars, and Chinese [t^h] was mapped to Japanese voiceless stop [t] as "fair" exemplar. According to PAM, as "fair" exemplars of Chinese unaspirated stops [p, t, k], the difficulty of distinguishing between these sounds and Japanese voiceless stops [p, t, k] was predicted to be "moderate". Similarly, the difficulty

of distinguishing between Chinese [p^h, k^h] and Japanese [p, k] for Japanese L2 learners was predicted to be "hard", and "moderate" for Chinese [t^h] and Japanese [t]. According to SLM, the production of a L2 sound eventually resembles the phonetic category established for this sound, and that the phonetic category of L2 sound established by L2 learners might be different from that of a monolingual, if this phonetic category is established based on different features or features weights than a monolingual. Acoustically Japanese voiceless stops have shorter VOT than Chinese aspirated voiceless stops and Japanese voiceless stops have longer VOT than Chinese unaspirated voiceless stops. The fact that the production of Chinese aspirated stops by Japanese speakers has shorter VOTs than the production of Chinese native speakers in this study confirmed this SLM prediction. The fact that the production of Chinese unaspirated stops by Japanese speakers has longer VOTs than the production of Chinese native speakers in this study also was consistent with SLM prediction.

For suprasegmental variables, Japanese L2 learners tended to be more variable in vowel durations (V%, ΔV, VarcoΔV, and nPVI), consonant durations (ΔC, VarcoΔC), and syllable duration (syllable SD). They appear to be non-native-like in tone and intonation patterns. In addition, Japanese L2 learners also tended to speak at a slower rate and had longer and more pauses, false starts and self-corrections. The theoretical framework of SLM and PAM mainly focus on segmental domains, and do not address prosodic aspects of second language speech learning. However, the findings in this study showing a large variability in rhythm, tone and intonation and fluency measures suggests that L2 learners may differ critically from native speakers in these areas. These findings indicated the importance of prosody in SLA (Second Language Acquisition) and would

also urge the theorists to incorporate prosodic domains in the development of theoretical framework so as to give a more accurate and fuller understanding of L2 learning phenomenon.

4.4. Rating study

This section first presents a rating experiment investigating how Japanese L2 learners' production of Mandarin Chinese sounds are rated in terms of perceived foreign accededness. The accentedness ratings provided by native listeners will be regressed on acoustic measurements obtained from previous production experiment in order to examine the relationship between acoustic characteristics of the L2 production and foreign accentedness.

4.4.1. Methods

4. 4.1.1. Participants

Eleven native Chinese listeners (5 female and 6 male) participated as raters in the foreign accent rating study, in which they examined accentedness of the speech samples produced by the L2 learners and native Chinese speakers. These raters were on average 30 years old (range: 28-36), and have never lived in an English-speaking country for more than 8 months. Most raters were from Jiangsu Province and one was from Sichuan Province. But the dialects they speak all belong to the northern dialect family of Chinese, which is very similar to Mandarin Chinese. None of the raters participated in the production task or the creation of the production prompts.

4.4.1.2. Stimuli

The speech samples of L2 learners and native speakers' productions of the 6 test sentences were used as stimuli and were presented to 11 native Chinese raters to examine

perceived accentedness of each production. The original speech samples were amplitude normalized to 75dB (Original speech samples). The original speech samples were also low-pass filtered to remove all energy components of the speech signal above 450 Hz and amplitude normalized to 75dB, resulting in another set of speech samples (Filtered speech samples). This treatment was implemented in order to eliminate some segmental information while retaining prosodic information (Trofimovich and Baker, 2006). The Original speech samples were used to obtain accentedness rating for speech that retains all acoustic information, both segmental and prosodic. The Filtered speech samples were used to obtain accentedness rating for speech that only retains prosodic information, so that it allowed us to examine the influence of prosody on perceived foreign accent in the absence of segmental information. Native speaker's speech samples were included in the rating materials to provide the anchor samples in the rating task. There were 396 stimuli in total (23 Japanese L2 learners and 10 Chinese native speakers x 6 sentences x 2 sets of Original and Filtered).

4.4.1.3. Procedure

In the rating task, 11 native Chinese raters listened to each utterance (a production of one of the 6 test sentences) and rated it on the degree of foreign accent. Each trial began with an auditory presentation of an utterance and the visual presentation of a visual analog scale (Urberg-Carlson, K., B. Munson, et al., 2009). Raters were then prompted to rate each utterance for degree of foreign accent by sliding the bar in the middle of the scale using a computer mouse. The leftmost point on the bar corresponded to "speech that sounded like that of a native Chinese speaker" and the rightmost point corresponded to "extremely strong foreign accent" as indicated on the screen. Raters were instructed

that they could drag the bar anywhere between those points according to their judgment of accentedness. An accent score between 0 and 100 was registered depending on where the bar was moved to between the two points (the leftmost point = 0, the rightmost point = 100). The raters had an option of listening to an utterance up to 5 times before making the final decision. All raters completed the rating task in approximately 45 minutes.

4.4.1.4. Analysis

Foreign accent ratings were Z-score normalized for each rater. A foreign accent rating score was obtained for each speaker (23 Japanese L2 learners and 10 native Chinese speakers) as the mean normalized accent ratings averaged across 6 sentences and 11 raters, with a greater value denoting a greater degree of foreign accent. As the summary of foreign accent rating scores show (Table 4.2), the scores of native Chinese speaker samples ranged below zero, indicating they were all rated less accented than the mean. Preliminary t-tests showed that native Chinese speakers' scores were significantly lower than the scores of Japanese L2 learners' ($p < .05$ for both ratings of original and filtered samples).

Given the preliminary results, foreign accent scores and acoustic measures were submitted to step-wise multiple regression analyses to explore the relationship between perceived foreign accent and acoustic characteristics of the speech samples. Two separate analyses were conducted. The first analysis examined the foreign accent rating scores of the Original speech sample as the dependent and all acoustic features measured as the predictors, including segmental features: F1 and F2 in vowels [i, ɨ], [y], [æ, a, ɑ], [u], [w] [e, ə, ɤ, o] and semivowels [w, j, ɥ]; closure duration and VOT in aspirated stops and unaspirated stops [p], [t], [k], [p^h], [t^h], [k^h], as well as prosodic features: C_ToBi,

ΔV, ΔC, V%, VarcoΔV, VarcoΔC, nPVI, articulation and speaking rate, pause duration and pause frequency, the number of false start and self correction. The second analysis examined the foreign accent rating scores of the Filtered speech sample as the dependent and only prosodic features were entered as the predictors. The separate analyses were conducted to put forward an effort to indirectly compare the influence of segmental and prosodic features on perceived foreign accent, as well as identifying different roles played by supersegmental features in terms of contributions to the perceived foreign accent.

Table 4.2. Summary of foreign accent rating scores

		Mean	SE	Minimum	Maximum
Japanese L2 Learners	Original samples	0.52	0.60	-1.25	1.28
	Filtered samples	0.50	0.58	-0.42	1.68
Chinese Native Speakers	Original samples	-1.20	0.23	-1.44	-1.44
	Filtered samples	-1.16	0.26	-0.66	-0.83

4.4.2. Results

4.4.2.1. Examining Both Segmental and Prosodic Factors

The best-fit model (Table 4.3) predicting foreign accent rating on Japanese L2 learners' and native Chinese speakers' original speech samples showed that C_Tobi Score, [k^h] closure duration, F1 of [ɑ] and ΔV contributed to foreign accent rating ($R = .995$, $p < .0001$). This model explained 99.5 % of the accent rating and indicated that Chinese Tobi Score was the most predictive factor (beta = -.766), followed by F1 in [a] (beta = -.360), ΔV (beta = .244) and [k^h] closure duration (beta =.147) (Table 4.4). Typically, variance inflation factor (VIF) greater than 10 is considered to merit further

investigation for multicollinearity and being redundant in the model (Neter et al., 1989, p. 409). The VIF of the factors retained in the model here were considerably smaller than 10.

The mean of the factors retained in the model are reported in Table 4.5 separately for Japanese L2 learners and Chinese native speakers. These results suggested that for Japanese L2 learners, non-native like tone and intonation pattern is the most important source contributing to the perception of foreign accent, followed by a long closure in stop [k^h], more front tongue position for the production of low back vowel [ɑ], and variations in vowel duration. These results indicate that L2 learners' non-native-like performance in tone and intonation was most robustly correlated with the accent rating. Following that, L2 learners' smaller value of F1 (higher tongue position) in [ɑ] and a greater variability in vowel duration were related to higher accent rating. Finally, longer duration of closure in aspirated stop [k^h] was also related to higher accent rating. In summary, almost all aspects that I examined, segmentals, tone and intonation, and rhythm, predict accent patterns of these learners. It is noteworthy that two prosodic features, tone and intonation score and ΔV (variation in vowel duration) were retained in the best-fit model. Initially, introducing filtered speech into the perception experiment was due to the consideration that the effects of suprasegmental features might be suppressed when segmental features are present at the same time. However, even in the presence of all the segmental features, several suprasegmental features were retained as important factors, and furthermore, tone and intonation score was the most influential feature in predicting perceived foreign accentedness.

Table 4.3. Model summary for original production

					Change Statistics				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change
1	.932	.869	.855	.303	.869	59.93	1	9	.000
2	.966	.933	.917	.230	.064	7.65	1	8	.024
3	.984	.967	.954	.171	.034	7.36	1	7	.030
4	.995	.990	.983	.103	.022	13.21	1	6	.011

Model 1: C_ToBiScore
Model 2: C_ToBiScore and [k^h] closure
Model 3: C_ToBiScore, [k^h] closure and F1 value of [ɑ]
Model 4: C_ToBiScore, [k^h] closure, F1 value of [ɑ] and ΔV

Table 4.4. Best model for original production

Best model for original production($p < .05$, $R^2 = .990$).

Significant Variable	β	*P*
C_ToBiScore	-.766	.000*
Closure of [k^h]	.147	.028*
F1 of [ɑ]	-.360	.001*
ΔV	.244	.011*
		* p< .05

Table 4.5. Means of the significant predictor variables

	Japanese L2 learners	Chinese Native Speakers
C_ToBiScore	24.83 (1.40)	27.33(0)
Closure of [k^h]	74.37 (12.94)	52.68 (10.09)
F1 of [ɑ]	1.38 (0.47)	1.48 (0.24)
ΔV	60.65(10.49)	46.97(9.66)

4.4.2.2. Examining Prosodic Factors Alone

The previous analysis revealed a strong influence of suprasegmental factors even in the presence of segmental factors. Accordingly, the following statistical analysis of filtered speech aimed to further identify a robustness among suprasegmental variables with respect to their contribution to the perception of foreign accent. Foreign accent rating for *filtered* production as the dependent variable with all suprasegmental variables

(C_Tobi Score, ΔV, ΔC, V%, VarcoΔV, VarcoΔC, nPVI, articulation and speaking rate, pause duration and pause frequency, the number of false start and self correction) as predictors were submitted to stepwise multiple regression. This best-fit model (Table 4.6) included Chinese Tobi Score, Speaking Rate and Self-correction Frequency as predictors successfully accounted for 89.5% of the variance in accent score of filtered production. The beta coefficients of the model (Table 4.7 and 4.8) indicated that Chinese Tobi Score (-.550) carried the most weight in influencing the perception of foreign accent, followed by Speaking Rate (.377), and Self-correction Frequency (.160). These results confirm that among suprasegmental features, failing to acquire native-like tone and intonation pattern is most important factor for accentedness. The results further indicated that in the absence of segmental issues, fluency measures are important factors that are related to accent rating.

4.5. General Discussion

Production study found, descriptively, that Japanese L2 learners' production differed from that of Chinese native speakers in terms of both segmental (higher F2 values in [ɨ], [u], [ə], lower F2 values in [y], [o], [a], shorter VOTs and longer closures in [ph, th, kh], longer VOTs and shorter closures in [p, t, k]) and suprasegmental domains

Table 4.6. Model summary for filtered production

					Change Statistics				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change
1	.914	.835	.830	.381	.835	159.70	1	31	.000
2	.934	.873	.864	.339	.038	8.97	1	30	.005
3	.946	.895	.884	.314	.022	6.16	1	29	.019

Model 1: C_ToBiScore
Model 2: C_ToBiScore and Speaking Rate
Model 3: C_ToBiScore, Speaking Rate and Self-correction Frequency(per sentence)

Table 4.7. Best model for filtered production

Best model for filtered production($p < .05$, $R^2 = .895$).

Significant Variable	β	*P*
C_ToBiScore	-.550	.000*
Speaking Rate	.377	.001*
Self-correction Freq	.160	.019*

* p< .05

Table 4.8. Means of the significant predictor variables.

	Japanese L2 learners	Chinese Native Speakers
C_ToBiScore	24.83 (1.40)	27.33(0)
Speaking rate	289.68(42.82)	276.07(40.95)
Self-correction Freq	0.1(0.18)	0(0)

(more variablility in V%, ΔV, VarcoΔV, nPVI, ΔC, VarcoΔC, and syllable SD, non-native-like tone and intonation pattern in C_Tobi Score, slower speech rate, longer and more pauses, false starts and self-corrections). The subsequent rating study investigated the relationship between the acoustic properties and accent ratings of the utterances obtained from native Chinese listeners. Regression analysis showed that not all features that were found different in L2 learners' production was retained in the best-fit prediction model of accent rating. Results showed that when both segmental and suprasegmental information was present in the stimuli, accent rating was influenced by two segmental factors, closure duration of aspirated stop [k^h] and the frequency of the first formant (F1) in [ɑ], as well as two suprasegmental factors, C_ToBi score, variance in the duration of vowels. Among them, tone and intonation pattern represented by C_ToBi score exerted the most important influence. This means that producing non-native-like lexical tones and intonation played the most important role in the perception of accentedness in Mandarin Chinese. F1 value of [ɑ] was the second most important contributing factor for foreign accent. The results indicated that if Japanese L2 learners produced low back vowel [ɑ]

without opening up their jaw wide enough, they were perceived as accented by native Chinese listeners. Variance in the duration of vowels (ΔV) was the third important predictor. L2 learners produced vowels with more variable durations and this was perceived as more accented. Closure in [k^h] was the fourth important factor. When the closure in [k] was longer, native listeners heard more degree of foreign accent.

C_ToBi score and linguistic rhythm (as characterized by variation in vowel duration) greatly influence the perception of foreign accent in Japanese learners' production of Mandarin Chinese. This result has the implication that Chinese as a tonal language, producing the correct lexical tones and intonational patterns is essential for the elimination of perception of foreign accent. Also, Japanese is defined as mora-timed language (e.g., Vance, 2008), in which the duration of an utterance is best characterized by the number of morae in that utterance (Port,1987). On the other hand, Mandarin Chinese is considered a syllable-timed language (Roach, 1982). As mentioned above, mora-timed Japanese tended to have larger %V due to its limited syllable types (Ramus et al, 1999). The results showing that Japanese L2 learners tended to produce much more variable vowel durations had proved that Japanese L2 learners have failed to adapt their L1 rhythmic features to Chinese ones in their production. At last, we noted that when both segmental and suprasegmental features were considered, none of the fluency measures affected the accent perception.

In the acoustic results the discrepancy in F1 value of [ɑ] between Japanese L2 learner group and Chinese native speaker group was not as large as other variables (Figure 4.2). This feature was, nonetheless, found to be one of the contributing factors in the best model. This might be due to the fact that acoustic discrepancies in L2 learners'

production do not necessarily translate equally to the perception of foreign accent. It is also possible that although the difference between L2 learners and native speakers' production in terms of F1 value of [ɑ] was not the largest, this amount of difference combined with other acoustic discrepancies was able to account the variance in perceived foreign accentedness most accurately. In addition, it should be reminded that this low back vowel [ɑ] only occurs in the phonetic environment of diphthongs ending in [u].

If we only focus on the contribution of suprasegmentals to the perception of foreign accent, as seen in the best model accounting for the perception rating scores for filtered production, without segmental information, Chinese ToBi Score, speech rate and self-correction frequency lead to accent perception. Both segmental and suprasegmental features accounted for 99.0% of the accentedness rating (Table 4.3), while suprasegmental features alone explained 89.5% of the accentedness rating (Table 4.5). In two best models for accent ratings for both filtered and unfiltered, Chinese ToBi Score was found to be the most substantial contributing factor for perceived foreign accent. From the best model account for original foreign accent rating containing both segmental and suprasegmental features, we have already seen that suprasegmentals carried more weights than segmentals. The effects of suprasegmentals on the perception of foreign accent can also be supported from the closeness between these two figures (99.0% vs. 89.5%).These results from the model with filtered production (Table 4.5) also suggest that a hierarchy of effects on foreign accent perception descending gradually from tone and intonation pattern to fluency, which is inconsistent with the finding of Trofimovich and Baker (2006) in their Korean accented English study. This inconsistence possibly

result from the fact that the target language Mandarin Chinese in current study is a tonal language while English in their study is not.

4.6. Conclusion

The results of this study, together with previous studies (e.g. Wayland, 1997; Trofimovich &Baker, 2006) on foreign accent, seem to suggest that acoustic variables contributing to the perception of accentedness are language-specific than universal across different languages. Wayland's (1997) study on English speakers learning Thai found that temporal variables such as VOT and vowel duration produced by L2 learners were not significantly different from that of native Thai speakers, while spectral variables such as vowel formants F1, F2 and fundamental frequency F0 were. In contrast, the current study found both temporal variable (k^h -closure) and spectral variable (F1 value of vowel [ɑ]) carried substantial perceptual weights in terms of foreign accentedness. Trofimovich &Baker (2006), found that fluency aspects (i.e., pause duration and speech rate) influenced perception of foreign accent more than rhythm or tonal aspects (i.e., stress timing, peak alignment) in English production of Korean learners. The current study found that speech rate was one of the important factors when listeners heard filtered speech (the same condition as Trofimovich and Baker (2006)); however, pause duration was not identified as a predictor. It is also possible, however, that the discrepancy in findings of these three foreign accent studies on three different target languages---Thai, English and Mandarin Chinese, comes from different methodologies and variables examined. For instance Wayland (1997) only examined a very limited number of Thai vowels and consonants, and Trofimovich &Baker (2006) only focused on suprasegmental variables. Further acoustic studies on foreign accent examining more language pairs are

still needed to give a conclusive answer on specificity vs. universality on foreign accent perception.

The findings of this study in both production and perception studies also provided important pedagogical implications that can serve as teaching guide for instructors of Japanese-speaking L2 learners of Mandarin Chinese. The results above confirming the substantial effects of suprasegmentals on the perception of foreign accent suggest that language instructors intoned to pay careful attention to teaching suprasegmental features of Chinese such as lexical tones, intonation and the importance of syllable-timing. Also, Japanese-speaking L2 learners of Mandarin Chinese also should try to speak with normal speed with as fewer pauses as possible if they want to sound like a native Chinese speaker. As far as teaching segmentals is concerned, the most effective teaching strategies to minimize foreign accent would be drawing distinction between Japanese [ɯ] and Chinese [u] and teaching L2 learners to pronounce Chinese [u] as a more back vowel rather than a central vowel. After distinguishing Japanese [ɯ] and Chinese [u], in addition, L2 instructors should also pay attention to L2 learners' pronunciation of Chinese stops to make sure they produce these stops with more aspiration compared to their Japanese counterparts.

In summary, this study has provided a deeper understanding of L2 foreign accent phenomenon through acoustic analysis of a less explored language pair in this field --- Chinese-Japanese. It was confirmed that acoustic sources for the perception of foreign accent come from both segmental and suprasegmental features of L2 speech. Suprasegmental features had more effects on the perceived foreign accentedness of L2 learners' speech than segmentals. However, due to the limited research efforts devoted to

acoustic examination of foreign accent phenomenon, more future acoustic research on foreign accent exploring more language pairs is also necessary in order to determine whether the perception of accentedness is language-specific or universal across different languages.

Chapter V

CONCLUSION

This dissertation conducted cross-linguistic investigation on how L2 learners acquire nonnative sounds in terms of both perception and production by looking at a less well-researched language pair---Japanese and Mandarin Chinese. This chapter summarizes the main findings in the following section with reference to two most influential theoretical models in second language learning---PAM and SLM. In addition, the pedagogical implication of the findings in second language acquisition will also be addressed.

5.1. Evaluation of Theoretical Models---PAM and SLM

PAM and SLM make theoretical predictions on categorical perception and production of nonnative sounds based on the similarities and dissimilarities between L1 and L2 sound categories. The findings in Chapter 2 and 3 with regard to the evaluation of PAM framework in the context of SLA are presented in 5.1.1 while findings in Chapter 4 in relation to SLM's hypothesis on the relationship between speech perception and production are presented in 5.1.2.

5.1.1. The Relationship Between Perception of L2 Sounds and L1 Phonology

The findings in categorical mapping and discrimination experiments in Chapter 2 only partially confirmed PAM's predictions on nonnative sound contrasts discriminability based on their assimilation types and perceived proximity to the most resembling L1 categories. As far as the general discriminability ranking of different assimilation types (SC, CG and SC) is concerned, since the discrimination experiment included fifteen vowel and consonant contrasts representative of four assimilation types: SC (Single-

Category) type, CG (Category-Goodness) type, UC (Uncategorized vs. Categorized) type, and UU (Both Uncategorizable) type, it enabled us to confirm the discrimination difficulty ranking of the first two types is: CG > SC, which was consistent with PAM's predictions as well findings in Best et al's (2001) study on English native speakers' perception of Zulu/Tigrinya consonant contrasts, but not in Harnsberger's study (2001).

However, the discrepancies observed in PAM's predictions of discriminability of a few sound contrasts indicated that the PAM does not always make correct predictions regarding second language speech learning. Specifically, PAM's predictions on UC type in which uncategorizable and categorizable sounds happen to share the same classified L1 categories are not instantiated by the findings in either discrimination study in Chapter 2 or training experiment in Chapter 3. Given the assumption that L1 and L2 sounds share the same phonological space by SLM, the perceived phonetic distances between these two sounds might shrink due to this overlap, leading to unexpected difficulty to discriminate. This finding resonated well with the findings in Guion et al's (2000) study on the language pair of English and Japanese, in which the discriminability of one sound contrast of UC type ([s]-[θ] contrast) was also found to be different from PAM's prediction. Discrepancies with other sound contrasts involving distinctive phonetic features specific to Mandarin Chinese (e.g. aspiration and retroflexion) made us speculate that perceived phonetic distances might not be the only factor influencing discriminability, it is also likely that specific phonetic or articulatory features can further facilitate or impede the discrimination of nonnative sounds.

The results in training experiment in Chapter 3 on one hand confirmed that L2 learners' perception can be improved by laboratory based training, which is encouraging

for L2 learners since significant performance improvement during this short period of training promises the high possibility of success in L2 acquisition in the future as language experiences increase. On the other hand, the sound contrast ([tʂʰ] vs. [ts]) that was predicted to be "easy" to discriminate by PAM was found to be surprisingly harder than sound contrast ([t] vs. [tʰ]) predicted to be "hard". This discrepancy further proved that the limitation of PAM in accurately accounting for discrimination difficulties for L2 learners by simply relying on the notion of perceived phonetic distances, and further reinforced our speculation that discriminability between sound contrasts also depended on other factors, such as certain distinctive phonetic features.

One possible explanation is that the feature of retroflexion is essentially much harder to perceive than aspiration for Japanese learners of Mandarin Chinese. This speculation can be supported by Japanese learners' familiarity with acoustic uses of aspiration (Voice Onset Time) since it has also been used in Japanese to distinguish voicing. Alternatively, the complex phonetic/articulatory profile of retroflex (Chang, Shih, & Allen, 2013) or the mix input resulted from retroflex and non-retroflex merger in Mandarin speech produced by Chinese from Southern regions (Zhang, 2012) could also provide supporting evidence for this argument. Another explanation is that these two features are possibly both hard to acquire initially. However, retroflexion is simply more resistant to perceptual changes brought by language experience than aspiration, since the same amount of Chinese instruction received by Japanese L2 learners was able to boost their identification accuracy to 92.89% at the very onset of the training, while their accuracy with retroflexion was still struggling at the level slightly better than chance (68.13%). Either case, the actual learning difficulty for certain nonnative sound contrasts

encountered by L2 learners are found to be much more complex than predicted by theoretical framework of PAM.

In addition, PAM is grounded in a direct realistic perspective of speech perception, which posits that perception process happens at the gestural level. Within this framework, listeners perceive speech sounds by extracting invariants about articulatory gestures from speech signals. Based on this notion, the difficulty in perceiving a specific feature such as aspiration should persist across all sounds having the feature of aspiration since the invariants extracted from speech signals about aspirated sounds should be same. However, the findings in category mapping and discrimination study presented in Chapter 2 showed otherwise. Based on perceived phonetic distances of Chinese sounds measured in terms of fit indices, the difficulty to discriminate between affricates pairs [ts^h] vs. [ts] and [tɕ^h] vs. [tɕ] which are only distinguished by aspiration were predicted by PAM to be "moderate". Similarly, stop pairs [k^h] vs.[k] and [p^h] vs. [p] were also predicted to be "moderate" to discriminate. However, discriminating between stop [t^h] vs. [t] was predicted to be "hard" although aspiration was the only phonetic feature that distinguishes this stop pair just as the other three contrast pairs.

The complex mechanism of L2 learners' perceptual learning was also noted by Best and her college (Best & Tyler, 2007) by stating that the perception of nonnative sounds by inexperienced listeners and L2 learners are different because L2 learners are pressured by the learning motivation to "re-phonologize" perception of the target sound contrasts while naive listeners are not. In addition, the perceptual learning of L2 learners' are speculated to be influenced by a range of factors, such as the specific phonetic features of the sound contrasts as shown in this project, or the linguistic learning

environment (whether the learning happens under FLA classroom or immersion SLA environment, whether the learners are simultaneous or early bilinguals), language experiences (which stage of learning they are positioned e.g. inexperienced vs. experienced learners) or age of L2 learning (whether they are adults learners or children). PAM, which makes predictions solely based on perceived phonetic distances between sound contrasts, is unable to accurately predict the perceptual learning difficulties L2 learners face. Accordingly, further research is necessary in order to find how discriminability between certain nonnative sound contrasts is influenced by these factors. Only after deciphering these mysteries can we be able to construct a theoretical speech perception model that accounts for the complexity of perceptual learning in L2 context.

5.1.2. The Relationship between L2 Perception and Production

As a speech perception model, PAM only focuses on the perception of nonnative sounds while Flege's SLM (1995) explicitly makes predictions on production in second language acquisition based on the notion that L2 learners' production of nonnative sounds are constrained by their perception. The findings in acoustic measurements of Japanese L2 learners' production of Mandarin Chinese in production study of Chapter 4 confirmed SLM hypothesis. Although it seems to be counterintuitive, SLM hypothesizes that nonnative sounds that are perceived to be more similar to L1 categories pose difficulty for L2 learning since equivalence classification prevents the establishment of a new category. The findings in categorical mapping study in Chapter 2 reported that certain sounds (e.g. [u]) were classified to Japanese categories as "good" exemplars. However, the acoustic measurements revealed that these sounds (e.g. [u]) in Japanese learners' production tended to be more divergent from Chinese native speakers' production. SLM's

hypothesis can easily explain this phenomenon. Japanese learners probably classified Chinese [u] as the equivalence of Japanese [ɯ], which prevents them from noticing the acoustic differences between these two sounds. The findings that Japanese L2 learners' production of Chinese [u] was acoustically resembled Japanese [ɯ] supported SLM's another hypothesis which predicts that the production of L2 sounds will eventually resemble the phonetic categories established for those sounds.

Another example instantiating SLM's hypothesis that L2 production is limited by their perception is the case of Chinese stops. The results in categorical mapping study of Chapter 2 indicated that Chinese aspirated and unaspirated stops were both assimilated into Japanese voiceless stops while aspirated stops were perceived to be better exemplars of this category. As for stop production, Japanese L2 learners tended to produce Chinese aspirated stops with much shorter VOTs and longer closure compared with Chinese native speakers. Classifying Chinese aspirated stops as equivalence to Japanese voiceless categories may have blocked the acoustic cues (longer VOT) of Chinese aspirated stops to be picked up by Japanese L2 learners, thus resulted in divergent production in L2 speech.

At last, SLM hypothesizes perceptual and production changes in L2 learning process across life span. It states that L2 learners' production will eventually resemble the categories established for them based on perception, suggesting that there might be a stage when a divergence between perception and production can be observed. If we relate empirical category mapping data in Chapter 2 and Chapter 3 together, we will discover that some sounds (e.g. Chinese central vowel [ə], and rounded front vowel [y]) were perceived to be quite different from L1 categories (Japanese vowel [o] and [i]), indicating

that L2 learners were able to discern phonetic differences between these two sounds and establish a new category accordingly. However, this does not guarantee immediate success in approximating production to that of native speakers. In fact, findings in production study of Chapter 4 revealed that there was great discrepancy between Japanese learners' production and native speakers' production of these two sounds (Chinese [ə] and [y]). This suggests that although production may be driven by perception as Flege (1995) assumes, the complex mechanism involved in production such as articulatory factors might results in asynchrony observed between production and perception.

5.2. Implication for Second Language Pedagogy

The findings in current dissertation have provided important pedagogical implications for L2 language instruction, especially for Chinese language instructors teaching Japanese L2 learners. Categorical mapping study in Chapter 2 provided the empirical data of the perceived phonetic distances between a comprehensive list of Chinese consonant and vowels and their closest Japanese categories. This information can serve as very useful reference when designing language exercises on certain sound contrasts by focusing on sound pairs that are predicted to be hard to discriminate and intentionally draw L2 learners' attention to these target sound contrasts.

Also, the findings in training experiment in Chapter 3 indicated that it is more difficult to identify sound contrasts distinguished by retroflexion than aspiration. Based on this finding, when these two features are taught in L2 classroom, it is advised that teaching syllabus can be designed in the way that efficiently allocates practice time based on the easiness/hardness of that sound feature. In this case, practicing on retroflexion

contrast needs to be weighted more heavily than aspiration. In addition, the effectiveness of laboratory training in Chapter 3 using HVPT directly informs us that it is necessary to provide more variable natural speech as language input in L2 classroom in order to facilitate perceptual improvement. For instance, teaching materials may include difficult segments and tones spoken by male and female speakers, in different position within words, spoken in different speech rate, and different types of sentences, since it is necessary for students to be able extract abstract prototypic cues resistant to talker differences and environment differences and to further form accurate and robust categorical representation of L2 sounds.

Although the communicative approach has gained great momentum in recent L2 pedagogical study, it is also necessary to give explicit instructions on how to discriminate and produce target sounds to L2 learners under classroom setting. Being able to accurately perceive L2 sounds is essential for meaningful and smooth communication. As Flege says, discerning different phonetic categories as different leads to setting up new categories, accordingly explicit instructions are helpful for students to notice differences between L2 sounds and their most resembling L1 categories, e.g. Chinese aspirated stops and Japanese voiceless stops. In addition, producing L2 sound in native-like fashion is also important since it was found that foreign accent can bring negative impression for the speaker since foreign accented speech makes the speaker sound less reliable and trustworthy (Lev-Ari & Keysar, 2010). The findings in foreign accent study in Chapter 4 provided concrete suggestions on how to help Japanese students learning Mandarin Chinese mitigate their foreign accent in speech production. The findings in Chapter 4 identified suprasegmental features (e.g. C_Tobi), which have been much less focused

than segmental features when teaching L2 learners pronunciation, as surprisingly the most important contributing factor to the perception of foreign accent. Accordingly, classroom instruction should work to raise students' awareness that acquiring Chinese tones and intonation are very important if they want to be understood clearly. For speech production of certain vowels and consonants, classroom instructions for Japanese L2 learners need to be explicit on how they should articulate certain Chinese sounds in order not to be perceived as foreign accented. For instance, they need to open up their jaw wide enough when producing Chinese low back vowel [ɑ] and at the same time shorten the closure of Chinese aspirated stop [k^h] in production.

APPENDIX A. READING LIST FOR CATEGORICAL MAPPING EXPERIMENT

素 su	入 ru	不 bu	酷 ku	聚 ju	应 ying	立 li	赖 lai	漏 lou	要 yao	亚 ya
数 shu	路 lu	铺 pu	顾 gu	去 qu	续 xu	路 lu	烙 lao	略 lue	院 yuan	由 you
租 zu	醋 cu	兔 tu	目 mu	付 fu		辣 la	烂 lan	龙 long	燕 yan	印 yin
住 zhu	触 chu	度 du	怒 nu	户 hu		绿 lv	浪 lang		样 yang	硬 ying
						乐 le	累 lei		夜 ye	问 wen
						落 luo	笨 ben		卧 wo	洼 wa
						二 er	愣 leng		月 yue	外 wai
									为 wei	万 wan
										望 wang

APPENDIX B. WORD LISTS USED IN TRAINING EXPERIMENT

Contrasts [t] vs. [t^h] [ts] vs. [tʂ]]

Pretest

他 ta	搭 da	1	栽 zai	摘 zhai	1
塔 ta	打 da	3	仔 zai	窄 zhai	3
踏 ta	大 da	4	再 zai	寨 zhai	4
胎 tai	呆 dai	1	资 zi	织 zhi	1
太 tai	带 dai	4	攥 zuan	转 zhuan	3
贪 tan	单 dan	1	字 zi	至 zhi	4
袒 tan	胆 dan	3	脏 zang	张 zhang	1
叹 tan	淡 dan	4	葬 zang	帐 zhang	4

Posttest

他 ta	搭 da	1	栽 zai	摘 zhai	1
塔 ta	打 da	3	仔 zai	窄 zhai	3
踏 ta	大 da	4	再 zai	寨 zhai	4
胎 tai	呆 dai	1	资 zi	织 zhi	1
太 tai	带 dai	4	攥 zuan	转 zhuan	3
贪 tan	单 dan	1	字 zi	至 zhi	4
袒 tan	胆 dan	3	脏 zang	张 zhang	1
叹 tan	淡 dan	4	葬 zang	帐 zhang	4

Training

他 ta	搭 da	1
塔 ta	打 da	3
踏 ta	大 da	4
胎 tai	呆 dai	1
太 tai	带 dai	4
贪 tan	单 dan	1
袒 tan	胆 dan	3
叹 tan	淡 dan	4
拖 tuo	多 duo	1
驼 tuo	夺 duo	2
妥 tuo	躲 duo	3
唾 tuo	剁 duo	4
涛 tao	刀 dao	1
讨 tao	倒 dao	3
套 tao	道 dao	4

栽 zai	摘 zhai	1
仔 zai	窄 zhai	3
再 zai	寨 zhai	4
资 zi	织 zhi	1
攥 zuan	转 zhuan	3
字 zi	至 zhi	4
脏 zang	张 zhang	1
葬 zang	帐 zhang	4
醉 zui	坠 zhui	4
作 zuo	桌 zhuo	1
琢 zuo	拙 zhuo	2
造 zao	罩 zhao	4
早 zao	找 zhao	3
遭 zao	招 zhao	1
子 zi	只 zhi	3

Gen1

听 ting	丁 ding	1
挺 ting	顶 ding	3
烫 tang	荡 dang	4
汤 tang	当 dang	1
躺 tang	挡 dang	3
透 tou	斗 dou	4
偷 tou	兜 dou	1
通 tong	东 dong	1
桶 tong	懂 dong	3
蹄 ti	笛 di	2

增 zeng	挣 zheng	1
尊 zun	谆 zhun	1
仄 ze	这 zhe	4
组 zu	主 zhu	3
簪 zan	沾 zhan	1
泽 ze	哲 zhe	2
赞 zan	站 zhan	4
邹 zou	周 zhou	1
走 zou	肘 zhou	3
攒 zan	展 zhan	3

Gen2

突 tu	嘟 du	1
图 tu	毒 du	2
土 tu	赌 du	3
兔 tu	度 du	4
踢 ti	低 di	1
舔 tian	点 dian	3
体 ti	抵 di	3
替 ti	第 di	4
贴 tie	爹 die	1
吞 tun	蹲 dun	1

宗 zong	中 zhong	1
钻 zuan	专 zhuan	1
租 zu	猪 zhu	1
足 zu	竹 zhu	2
咋 za	眨 zha	3
怎 zen	缜 zhen	3
谮 zen	阵 zhen	4
匝 za	扎 zha	1
揍 zou	咒 zhou	4
紫 zi	纸 zhi	3

APPENDIX C. PROMPTS FOR THE DELAYED REPETITION TASK

(1) Question: *Na shi shen me za zhi?* "What kind of magazine is that?"
Response: *Na shi ying wen za zhi.* "That is English magazine."

(2) Question: *Ni qu na li?* "Where are you going?"
Response: *Wo qu mai ping guo bi ji ber dian nao.* "I go out to buy an Apple laptop."

(3) Question: *Ni men he dianr shen me?* "What would you like to drink?"
Response: *Wo he ka fei. Ta he cha.*. "Coffee for me, and tea for him."

(4) Question: *Ni mai shen me dong xi?* "What do you want to buy?"
Response: *Wo mai san ge ben zi, yi jian yi fu he yi pan ci dai.* "I want to buy three notebooks, one piece of clothes, and one cassette tape."

(5) Question: *Tian fang he zhang dong ye shi ri ben liu xue sheng ma?* "Are Tian Fang and Zhang Dong Japanese students?"
Response: *Bu. Ta men bu shi ri ben liu xue sheng. Ta men dou shi zhong guo xue sheng.*
"No, they are not Japanese students. They are Chinese students."

(6) Question: *Ima naji ga ichiban hoshiidesu ka?* "Do you think learning Chinese is hard?"
Response: *Wo jue de ting shuo bi jiao rong yi, du xie hen nan.*. "I think listening and speaking is relatively easy while reading and speaking is hard."

REFERENCES CITED

Anderson-Hsieh, J., Johnson, R., & Koehler, K. (1992). The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentais, prosody, and syllable structure. *Language Learning, 42*(4), 529-555.

Aslin, R. N., & Pisoni, D. B. (1980). Some developmental processes in speech perception. *Child phonology, 2*, 67-96.

Best, C. T. (1995). Chapter 6: A Direct Realist View of Cross-Language Speech Perception. *Speech perception and linguistic experience: Issues in cross-language research*, 171-204.

Best, C. T., McRoberts, G. W., & Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *The Journal of the Acoustical Society of America, 109*(2), 775-794.

Best, C. T., & Strange, W. (1992). Effects of phonological and phonetic factors on cross-language perception of approximants. *Journal of Phonetics, 20*(3), 305-330.

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. *Language experience in second language speech learning: In honor of James Emil Flege, 1334*.

Boersma, P. (2002). Praat, a system for doing phonetics by computer. *Glot international, 5*(9/10), 341-345.

Boersma, P., & Weenink, D. (2005). Praat software (version 5.2. 01): Amsterdam, Universidad de Amsterdam. http://www. fon. hum. uva. nl/praat.[11/11/2012].

Bradlow, A. R. (2008). Training non-native language sound patterns: lessons from training Japanese adults on the English. *Phonol. Second Lang. Acquis, 36*, 287-308.

Bradlow, A. R., Akahane-Yamada, R., Pisoni, D. B., & Tohkura, Y. i. (1999). Training Japanese listeners to identify English/r/and/l: Long-term retention of learning in perception and production. *Perception & Psychophysics, 61*(5), 977-985.

Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., & Tohkura, Y. i. (1997). Training Japanese listeners to identify English/r/and/l: IV. Some effects of perceptual learning on speech production. *The Journal of the Acoustical Society of America, 101*(4), 2299-2310.

Chan, A. Y. (2007). The acquisition of English word-final consonants by Cantonese ESL learners in Hong Kong. *The Canadian Journal of Linguistics/La revue canadienne de linguistique, 52*(3), 231-253.

Chang, Y., Shih, C., & Allen, J. (2013). *Variability in cross-dialect perception of the Mandarin alveolar–retroflex contrast.* Paper presented at the Proceedings of the International Conference on Phonetics of the Languages in China.

Chao, K.-Y., & Chen, L.-m. (2008). A cross-linguistic study of voice onset time in stop consonant productions. *Computational Linguistics and Chinese Language Processing, 13*(2), 215-232.

Dellwo, V. (2006). Rhythm and speech rate: A variation coefficient forΔ C. *Language and language-processing*, 231-241.

Eckman, F. R. (1977). Markedness and the contrastive analysis hypothesis. *Language Learning, 27*(2), 315-330.

Elliott, A. R. (1995). Foreign language phonology: Field independence, attitude, and the success of formal instruction in Spanish pronunciation. *The Modern Language Journal, 79*(4), 530-542.

Flege, J. E. (1991). Perception and production: The relevance of phonetic input to L2 phonological learning. *Crosscurrents in second language acquisition and linguistic theories, 2*, 249-289.

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. *Speech perception and linguistic experience: Issues in cross-language research*, 233-277.

Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds "L" and "R". *Neuropsychologia, 9*(3), 317-323.

Grabe, E., & Low, E. L. (2002). Durational variability in speech and the rhythm class hypothesis. *Papers in laboratory phonology, 7*(515-546).

Guion, S. G., Flege, J. E., Akahane-Yamada, R., & Pruitt, J. C. (2000). An investigation of current models of second language speech perception: The case of Japanese adults' perception of English consonants. *The Journal of the Acoustical Society of America, 107*(5), 2711-2724.

Harnsberger, J. D. (2001). On the relationship between identification and discrimination of non-native nasal consonants. *The Journal of the Acoustical Society of America, 110*(1), 489-503.

Hirata, Y., Whitehurst, E., & Cullings, E. (2007). Training native English speakers to identify Japanese vowel length contrast with sentences at varied speaking ratesa). *The Journal of the Acoustical Society of America, 121*(6), 3837-3845.

Hoshino, A., & Yasuda, A. (2006). *Evaluation of aspiration sound of Chinese labial and alveolar diphthong uttered by Japanese students using voice onset time and breathing power.* Paper presented at the Proceeding of ISCSLP.

Idemaru, K., & Guion-Anderson, S. (2010). Relational timing in the production and perception of Japanese singleton and geminate stops. *Phonetica, 67*(1-2), 25-46.

Ingram, J. C., & Park, S.-G. (1998). Language, context, and speaker effects in the identification and discrimination of English/r/and/l/by Japanese and Korean listeners. *The Journal of the Acoustical Society of America, 103*(2), 1161-1174.

Iverson, P., Hazan, V., & Bannister, K. (2005). Phonetic training with acoustic cue manipulations: A comparison of methods for teaching English/r/-/l/to Japanese adults. *The Journal of the Acoustical Society of America, 118*(5), 3267-3278.

Jin, L. (2008). *Markedness and second language acquisition of word order in Mandarin Chinese.* Paper presented at the Proceedings of the 20th North American Conference on Chinese Linguistics.

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (2006). Linguistic experience alters phonetic perception in infants by 6 months of age. *Foundations of Pediatric Audiology*, 71.

Labrune, L. (2012). *The phonology of Japanese*: Oxford University Press.

Lado, R. (1957). Linguistics Across Cultures: Applied Linguistics for Language Teachers.

Lee, W.-S., & Zee, E. (2003). Standard Chinese(Beijing). *Journal of the International Phonetic Association, 33*(1), 109-112.

Lev-Ari, S., & Keysar, B. (2010). Why don't we believe non-native speakers? The influence of accent on credibility. *Journal of Experimental Social Psychology, 46*(6), 1093-1096.

Lin, Y.-H. (2007). *The Sounds of Chinese with Audio CD* (Vol. 1): Cambridge University Press.

Lively, S. E., Logan, J. S., & Pisoni, D. B. (1993). Training Japanese listeners to identify English/r/and/l/. II: The role of phonetic environment and talker variability in learning new perceptual categories. *The Journal of the Acoustical Society of America, 94*(3), 1242-1255.

Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English/r/and/l: A first report. *The Journal of the Acoustical Society of America, 89*(2), 874-886.

Long, M. H. (1990). Maturational constraints on language development. *Studies in second language acquisition, 12*(03), 251-285.

MacKain, K. S., Best, C. T., & Strange, W. (1981). Categorical perception of English/r/and/l/by Japanese bilinguals. *Applied Psycholinguistics, 2*(04), 369-390.

Mattock, K., & Burnham, D. (2006). Chinese and English infants' tone perception: Evidence for perceptual reorganization. *Infancy, 10*(3), 241-265.

McClaskey, C. L., Pisoni, D. B., & Carrell, T. D. (1983). Transfer of training of a new linguistic contrast in voicing. *Perception & Psychophysics, 34*(4), 323-330.

Missaglia, F. (1999). *Contrastive prosody in SLA: An empirical study with adult Italian learners of German.* Paper presented at the Proceedings of the 14th International Congress of Phonetic Sciences.

Miyawaki, K., Jenkins, J. J., Strange, W., Liberman, A. M., Verbrugge, R., & Fujimura, O. (1975). An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. *Perception & Psychophysics, 18*(5), 331-340.

Mok, P. (2009). On the syllable-timing of Cantonese and Beijing Mandarin. *Chinese Journal of Phonetics, 2*, 148-154.

Moyer, A. (1999). Ultimate attainment in L2 phonology. *Studies in second language acquisition, 21*(01), 81-108.

Munro, M. J. (1995). Nonsegmental factors in foreign accent. *Studies in second language acquisition, 17*(01), 17-34.

Nearey, T. M. (1978). *Phonetic feature systems for vowels* (Vol. 77): Indiana University Linguistics Club.

Okada, H. (1991). Japanese. *Journal of the International Phonetic Association, 21*(02), 94-96.

Oyama, S. (1976). A sensitive period for the acquisition of a nonnative phonological system. *Journal of Psycholinguistic Research, 5*(3), 261-283.

Patkowski, M. S. (1990). Age and accent in a second language: A reply to James Emil Flege. *Applied linguistics, 11*(1), 73-89.

Piske, T., MacKay, I. R., & Flege, J. E. (2001). Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics, 29*(2), 191-215.

Pisoni, D. B. (1992). *Some comments on invariance, variability and perceptual normalization in speech perception.* Paper presented at the Second International Conference on Spoken Language Processing.

Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of experimental psychology, 77*(3p1), 353.
Ramus, F., Nespor, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition, 73*(3), 265-292.
Riney, T. J., Takagi, N., Ota, K., & Uchida, Y. (2007). The intermediate degree of VOT in Japanese initial voiceless stops. *Journal of Phonetics, 35*(3), 439-443.
Schmidt, A. M. (2007). Cross-language consonant identification: English and Korean. *Language experience in second language speech learning: In honor of James Emil Flege, 17*, 185.
Scovel, T. (1988). *A time to speak: A psycholinguistic inquiry into the critical period for human speech*: Newbury House Publishers.
Sheldon, A., & Strange, W. (1982). The acquisition of/r/and/l/by Japanese learners of English: Evidence that speech production can precede speech perception. *Applied Psycholinguistics, 3*(03), 243-261.
Shimizu, K. (1989). A cross-language study of voicing contrasts of stops. *America, 66*(4), 1001-1017.
Smith, J., & Kochetov, A. (2009). Categorization of non-native liquid contrasts by Cantonese, Japanese, Korean, and Mandarin listeners. *Toronto Working Papers in Linguistics, 34*.
Snndgrass, J. G., Levy—Berger, G., & Haydon, M. (1985). Human experimental psychology.
Strange, W., & Dittmann, S. (1984). Effects of discrimination training on the perception of/rl/by Japanese adults learning English. *Perception & Psychophysics, 36*(2), 131-145.
Suter, R. W. (1976). Predictors of pronunciation accuracy in second language learning1. *Language Learning, 26*(2), 233-253.
Svantesson, J.-O. (1984). Vowels and diphthongs in standard Chinese. *Working Papers (Lund University, Department of Linguistics), 27*, 209-235.
Tahta, S., Wood, M., & Loewenthal, K. (1981). Foreign accents: Factors relating to transfer of accent from the first language to a second language. *Language and Speech, 24*(3), 265-272.
Tees, R. C., & Werker, J. F. (1984). Perceptual flexibility: maintenance or recovery of the ability to discriminate non-native speech sounds. *Canadian Journal of Psychology/Revue canadienne de psychologie, 38*(4), 579.
Thomas, E. R., & Kendall, T. (2007). NORM: The vowel normalization and plotting suite. *Online Reso urce: http://ncslaap. lib. ncsu. edu/tools/norm*.
Thomas Erik, R. (2011). Sociophonetics: An introduction. *Basingstoke and New York: Palgrave Macmillan*.
Thompson, I. (1991). Foreign accents revisited: The English pronunciation of Russian immigrants. *Language Learning, 41*(2), 177-204.
Trehub, S. E. (1976). The discrimination of foreign speech contrasts by infants and adults. *Child development*, 466-472.
Trofimovich, P., & Baker, W. (2006). Learning second language suprasegmentals: Effect of L2 experience on prosody and fluency characteristics of L2 speech. *Studies in second language acquisition, 28*(01), 1-30.
Trubetzkoy, N. S. (1969). Principles of phonology.

Tsao, F.-M., Liu, H.-M., & Kuhl, P. K. (2006). Perception of native and non-native affricate-fricative contrasts: Cross-language tests on adults and infants. *The Journal of the Acoustical Society of America, 120*(4), 2285-2294.

Tsujimura, N. (2013). *An introduction to Japanese linguistics*: John Wiley & Sons.

Vance, T. J. (2008). *The sounds of Japanese with audio CD*: Cambridge University Press.

Venditti, J. J. (2006). The J_ToBI Model of Japanese. *Prosodic typology: The phonology of intonation and phrasing, 1*, 172.

Wang, Y., Jongman, A., & Sereno, J. A. (2003). Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training. *The Journal of the Acoustical Society of America, 113*(2), 1033-1043.

Wang, Y., Spence, M. M., Jongman, A., & Sereno, J. A. (1999). Training American listeners to perceive Mandarin tones. *The Journal of the Acoustical Society of America, 106*(6), 3649-3658.

Wayland, R. (1997). Non-native production of Thai: Acoustic measurements and accentedness ratings. *Applied linguistics, 18*(3), 345-373.

Wayland, R. P. (2007). The relationship between identification and discrimination in cross-language perception. *Language experience in second language speech learning: In honor of James Emil Flege, 17*, 201.

Werker, J. F. (1989). Becoming a native listener. *American Scientist, 77*(1), 54-59.

Werker, J. F., Gilbert, J. H., Humphrey, K., & Tees, R. C. (1981). Developmental aspects of cross-language speech perception. *Child development*, 349-355.

Werker, J. F., & Logan, J. S. (1985). Cross-language evidence for three factors in speech perception. *Perception & Psychophysics, 37*(1), 35-44.

Werker, J. F., & Tees, R. C. (1983). Developmental changes across childhood in the perception of non-native speech sounds. *Canadian Journal of Psychology/Revue canadienne de psychologie, 37*(2), 278.

Werker, J. F., & Tees, R. C. (1984a). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant behavior and development, 7*(1), 49-63.

Werker, J. F., & Tees, R. C. (1984b). Phonemic and phonetic factors in adult cross-language speech perception. *The Journal of the Acoustical Society of America, 75*(6), 1866-1878.

White, L., & Mattys, S. L. (2007). Calibrating rhythm: First language and second language studies. *Journal of Phonetics, 35*(4), 501-522.

Wong, J. W. S. (2012). Training the Perception and Production of English /e/ and/æ/ of Cantonese ESL Learners: A Comparison of Low vs. High Variability Phonetic Training.

Zhu, L. (2012). *Retroflex and non-retroflex merger in Shanghai accented Mandarin.* University of Washington.