Mandarin Retroflex Sounds Perceived by Non-native Speakers*

Shu-chen Ou ** Zhe-chen Guo***

Abstract

This study investigates two articulatory variants of the Mandarin retroflex as perceived by non-native speakers who learned Mandarin Chinese as a second/foreign language (L2). A review of Mandarin teaching materials and previous studies reveals that the following two gestures produce Mandarin retroflexes: “curling up” of the tongue and “bunching up” of the tongue. The focal question is whether these two gestures affect non-native speakers’ perceptions differently when they are distinguished from their alveolar (non-retroflex) counterparts. To address this question, an ABX discrimination task involving Taiwan Mandarin speakers (as a control group) and L2 Mandarin learners (as experimental groups) was implemented. Data were statistically analyzed for Taiwan Mandarin, Japanese, English, and French participants, resulting in the

* This research was sponsored by the National Science Council (project number: NSC 102-2815-C-110-010-H). We appreciate the comments of two anonymous reviewers regarding this paper.
** 國立中山大學外國語文學系助理教授。
*** 國立中山大學外國語文學系碩士生。
◎ 本文收稿日期：2014 年 5 月 12 日；審查通過日期：2014 年 9 月 22 日。

following two notable findings: i) the two types of retroflexes were not perceived differently by the non-native speakers when paired with their alveolar counterparts, and ii) the mean accuracy rate of the Japanese group was significantly lower than those of the English and French groups. One explanation for the Japanese participants' poor performance is that they are able to detect the phonetic differences between Mandarin post-alveolars and Japanese alveolo-palatals. This explanation implies that non-native speakers may be able to distinguish the two retroflex variants if their attention is drawn to the phonetic details. These findings have several implications for Mandarin pedagogy.

Keywords: Mandarin retroflex, articulatory gesture, second-language acquisition, non-native speakers' perception, Mandarin teaching/learning

非母語人士對華語捲舌音的感知

歐淑珍、郭哲宸

摘要

華語文教材與前人之研究顯示，「捲起」與「拱起」舌位皆可做為華語捲舌音發音方式，而此兩種方式在聽覺上對非母語人士而言是否不同仍待解答。本研究探討此問題，採用 ABX 區分實驗，以華語母語人士作控制組，並招募以華語為二語的學習者作為實驗組（英語、法語、日語共三組），隨後統計分析各組之實驗數據。有兩項結果值得注意：一、在各自與對應的齒齦音比較時，除母語組以外，兩種發音方式的捲舌音對非母與人士在聽辨上幾無差別；二、日語組的反應正確率顯著低於其他兩組。當假設日語母語人士能察覺華語齒齦後音與日語齦齶音之間的語音差異，則可解釋為何該組表現較差。而此假設更暗示，如非母語人士將注意力放在語音細節上時，其可能能夠分辨兩種捲舌音發音方式之間的不同。最後，研究結果為華語教學提供些許建議。

關鍵詞：華語捲舌音、發音方式、二語習得、非母語人士之聽辨、華語教學

Mandarin Retroflex Sounds Perceived by Non-native Speakers

Shu-chen Ou Zhe-chen Guo

1. Introduction

This study explores the effects of two types of Mandarin retroflex articulation on non-native speakers' perception of retroflex vs. alveolar contrasts. In Mandarin Chinese, the alveolar consonants /ts/, /tsʰ/, and /s/ are contrastive with their retroflex counterparts /ʈʂ/, /ʈʂʰ/, and /ʂ/. In particular, the pronunciation of the retroflex consonants has been cited as challenging for non-native learners. To overcome this challenge, a number of pedagogical materials have offered explicit articulatory descriptions of how the tongue should be positioned to produce those sounds. Most of the materials indicate that the sounds involve a gesture in which the tongue tip is raised and sometimes curled back against the back of the alveolar ridge (e.g., Ye, 1994; Wang, 2005; Lee, Dai, & Guo, 1995; Lu, 2006). However, these descriptions may not be relevant for all native Mandarin speakers. According to Ladefoged and Wu (1984), the retroflex sounds do not involve a curled tongue shape; instead, they are produced by lowering the tip of the tongue and bunching up the tongue body. These discrepant views merit

investigation given that non-native learners (or their instructors) may rely on pedagogical materials to produce (or help their students produce) /ʈʂ/, /ʈʂʰ/, and /ʂ/. Thus, the current study examines whether the two variants of the retroflex sound different to non-native speakers.

To examine this question, an experiment was designed to test predictions for non-native speakers with different first-language (L1) backgrounds. Three groups of non-native Mandarin learners were selected for the experimental group, sorted by their L1 as follows: Japanese, English, and French. Before the experiment, the two gestural types of retroflexes were compared to predict how they would be perceived by the non-native speakers. The predictions invoke Best's (1995) Perceptual Assimilation Model (PAM), which is a model of phonological perception based on the categorization of L2 phones into L1 phonemes. The phonological systems of the three languages were then examined, and a general PAM prediction suggested that the learners would classify the two variants of the retroflexes into the same category if alveolar vs. post-alveolar contrasts existed for fricatives or affricates in their native languages. To test this prediction, the experiment was conducted.

Another motivation to answer the research questions comes from the hope that the results of the current study would facilitate Mandarin instructors' teaching and non-native speakers' learning of the retroflex articulation. Instructors often confront problems, such as ineffective communication among their beginning learners, when the learners come from various L1 backgrounds but all speak Mandarin as a lingua franca. One way to address this issue is to consider the case of English as an International Language (EIL) and draw on the EIL framework from Jenkins (2000), which emphasizes intelligibility among non-native speakers.

If this concept is applied to a Mandarin pedagogical context, intelligibility becomes the focus of the issue and the current study may be able to offer several suggestions for improvement. For example, if the use of a particular type of retroflex leads to better perceptual distinction between the retroflexes and alveolars, then instructors should be encouraged to teach this type of retroflex on the ground that it promotes intelligibility.

Thus, the current study has a twofold purpose: 1) to investigate whether differences between the two gestural types of retroflexes affect non-native speakers' perception of retroflex vs. alveolar contrasts and 2) to provide suggestions to improve Mandarin pedagogy.

2. Literature Review

2.1 "Curled-up" vs. "bunched-up" retroflexes

A popular view held by native speakers of Mandarin Chinese is that its retroflex sounds—namely /ʈʂ/, /ʈʂʰ/, and /ʂ/—are produced by raising the tongue tip upwards and then backwards. This belief is evident in the majority of studies and teaching/learning materials (e.g., Ye, 1994; Wang, 2005; Lee, *et al*., 1995; Lu, 2006), which occasionally provide diagrams to illustrate the articulation of these sounds. Figures 1 and 2 are examples from Ye (1994) and from a book by the editorial committee at the National Taiwan Normal University (2007), respectively. These diagrams show that the tongue tip is raised to approach the region between the hinder part of the alveolar ridge and the forward part of the palate. These retroflex

sounds belong to the post-alveolar category[1] and have been described in detail in previous research. For example, Chao (1948, 1968, as cited in Lin, 2007) considers the constriction at the post-alveolar region as formed by the underside of the tongue tip, whereas Lee and Zee (2003, as cited in Lin, 2007) and Zee (2003, as cited in Lin, 2007) contend that the upper side is actually utilized. Regardless of which tongue side contacts the back of the alveolar ridge, most studies and teaching/learning materials posit that the sounds in question are produced with a curled-up tongue tip. In the present study, these retroflexes are termed curled-up (hereafter, CU) retroflexes. They can be distinguished from another type of retroflex in which the tongue tip is neither raised nor curled.

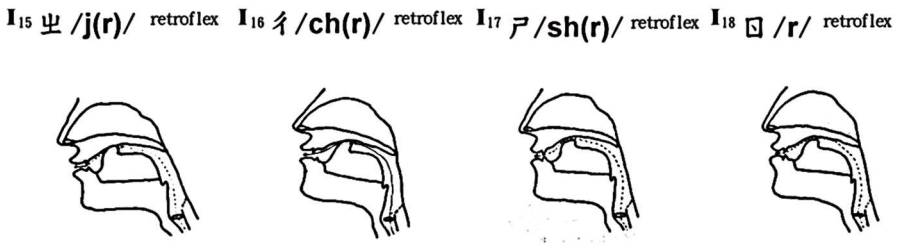

Figure 1 From left to right, articulations of /ʈʂ/, /ʈʂʰ/, /ʂ/, and /ɻ/[2] (Ye, 1994: 9-10)

---

1 It must be noted that "post-alveolar" is sometimes a cover term for sounds with places of articulation in this region. Retroflexes, palate-alveolars, and alveolo-palatals are post-alveolars that differ in their degree of palatalization. For this reason, some scholars (e.g., Lin, 2007) use "post-alveolar" exclusively for retroflex sounds, such as [ʂ] in Mandarin.

2 Transcribed in Hanyu Pinyin as *r*, /ɻ/ (a voiced post-alveolar approximant) tends to vary with the voiced retroflex fricative [ʐ] and is traditionally classified as a retroflex (Lin, 2007). Nevertheless, because this action suggests that /ɻ/ may still be realized as an approximant, the current research will limit its discussion by focusing on the voiceless fricative and affricates.

Figure 2 From left to right, articulations of /ʈʂ/, /ʈʂʰ/, /ʂ/, and /ɻ/
(The Editorial Board of Chinese Phonetics Pedagogical Materials at the National Taiwan Normal University, 2007: 136-142)

Ladefoged and Wu (1984) note another articulatory variant of the retroflex while investigating how /ʈʂ/, /ʈʂʰ/, and /ʂ/ are produced by three Beijing Mandarin speakers. Figure 3 presents the speakers' articulatory diagrams: instead of raising the tip of the tongue, the speakers lower the tongue and move its blade upwards to contact the back of the alveolar ridge and slightly hump the tongue dorsum. A comparison of Figure 3 with Figures 1 and 2 reveals the differences in the gestures employed by the Beijing Mandarin and CU speakers. Ladefoged and Maddieson (1996) further discuss the speakers' [ʂ] as a laminal flat post-alveolar sibilant.[3] Given that the differences between /ʈʂ/, /ʈʂʰ/, and /ʂ/ are in aspiration and manner of articulation, one assumption is that the retroflexes of the Beijing Mandarin speakers in Ladefoged and Wu (1984) share one feature in that they are all laminal. In the current context, laminal retroflexes are referred to as the "bunched-up" (BU) type.

[3] Ladefoged and Maddieson (1996) regard the term "retroflex" as an inappropriate articulatory description of *sh*, which is the Pinyin transcription of the voiceless retroflex fricative in Mandarin. Therefore, they transcribe this as [ʂ], an *s* with an underdot, and reserve the symbol [ʂ] for the sub-apical retroflex in Toda. To be consistent with the conventional transcriptions of Mandarin retroflexes, [ʂ] is used in this paper.

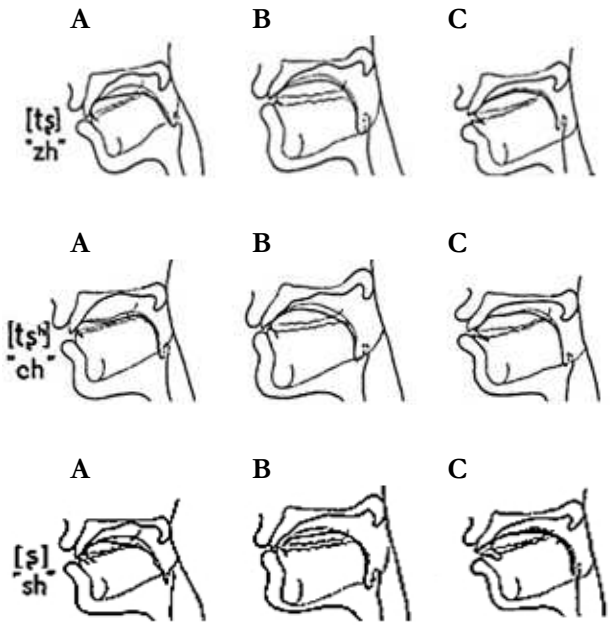

Figure 3 Articulatory diagrams of three speakers' [tʂ], [tʂʰ], and [ʂ] (Ladefoged & Wu, 1984: 269-273)

Interestingly, a similar phenomenon has been observed for the approximant /r/[4] in American English, which may be accounted for from an auditory-acoustic perspective. Studies have indicated that English speakers demonstrate articulatory variability in their production of the approximant and the variations can be categorized into two main types: "retroflex-*r*" and "bunched-*r*" (Ball & Rahilly, 1999; Guenther, *et al*., 1999; Bickford & Floyd, 2006; Zhou, *et al.*, 2008). For a comparison of these two types, Figure 4 presents diagrams comparing these two

[4] This symbol is used in some books (e.g., Ladefoged & Johnson, 2011), although this English approximant is transcribed as [ɹ] in the International Phonetics Alphabet (IPA).

types, as cited by Ball and Rahilly (1999).

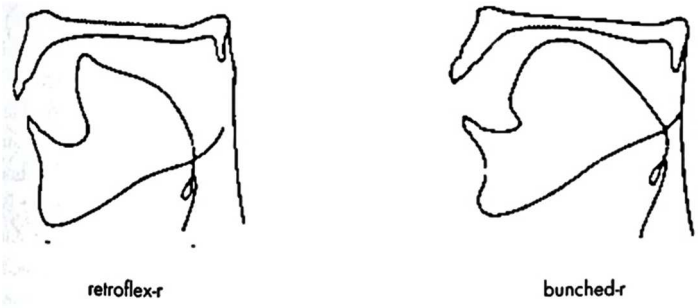

Figure 4 Retroflex-*r* and bunched-*r* (Ball & Rahilly, 1999: 89)

At this point, the question arises as to why there are these variants in the approximant /r/. One possible explanation would be related to their being non-articulatory, for example, the retroflex-*r* and bunched-*r* are both associated with lowering the third and fourth formants (Delattre, 1951, as cited in Lindau, 1978). Thus, speakers are able to vary their gestures as long as they achieve the same acoustic effect.

With regard to Mandarin Chinese, the question arises whether CU and BU variants also share common properties in the acoustic and auditory domains. Studies (e.g., Jakobson, Fant, & Halle, 1951; Ladefoged, 2001: 212–214, as cited in Wan & Jaeger, 2003) suggest that retroflexion typically results in the lowering of the first three formants, particularly the third one. This tendency was observed for both retroflex variants in a post-hoc acoustic analysis of the CU and BU retroflex sound tokens used in a pilot study by Guo and Ou (2014). Nevertheless, differences between the two variants were revealed when another technique for

acoustic measurement was employed. Cheng (2006, 2009) applies moment analysis to the investigation of Mandarin retroflexes and indicates that stronger retroflexion is associated with lower first moment (M1), which is an effective parameter for acoustically distinguishing between retroflexes and alveolars. The moment analyses have also been applied to measure differences in M1 between the CU and BU tokens utilized by Guo and Ou: The results showed that the average M1 values of the CU and BU retroflex consonants were 2.85 kHz and 3.37 kHz, respectively, suggesting that the retroflexes produced during the two types of gestures were acoustically different.

Although the CU and BU retroflexes should be perceived by native and non-native Mandarin speakers as allophones of the same retroflex categories so that both will be distinguished from alveolars, they actually differ articulatorily and acoustically. However, it is still unclear how their articulatory and acoustic differences translate into differences in the auditory domain. Therefore, the current research investigates the following question: Will the CU and BU retroflexes be perceived differently in Mandarin retroflex vs. alveolar contrasts?

2.2 Perceptual Assimilation Model (Best, 1995)

The PAM proposed by Best may help answer this question, as it assumes that one clue to perceptual similarity is perceived gestural information. The PAM adopts a direct realist perspective in which listeners detect articulatory gestures in speech signals. This premise serves as the basis for the similarity perception of non-native phones in the model, as follows:

> "Because the universal phonetic domain and native phonological space are defined by the spatial layout of the vocal tract and the dynamics characteristics of articulatory gestures, those distal properties provide the dimensions within which similarity is judged" (Best, 1995: 193).

The articulatory gesture of a segment is specified by, for example, the articulators employed and the location of the constriction along the vocal tract. When information about these gestures is perceived, the similarity of the segment to some native phonemes is evaluated. If there are two phones in a non-native contrast, then the manner in which they are categorized into native phonemes determines the extent to which they are similar or different for the listener. For example, in Best, McRoberts, and Goodell (2001), American English listeners are predicted to categorize voiceless vs. voiced lateral fricatives (/ɬ/ vs. /ɮ/) in Zulu into two different categories; that is, they should perceive the Zulu contrast in a Two-Category (TC) fashion. They show this discrimination because there are contrastive phones in American English that are articulatorily similar to the Zulu consonants but different in voicing.

Applying the PAM from Best (1995) to the variation of the Mandarin retroflex generates an overall prediction for non-native speakers. The aforementioned comparison of the CU and BU retroflexes indicated that they differ in terms of which part of the tongue is used: The former is apical, whereas the latter is laminal. Since Mandarin does not assign phonological functions to the apical vs. laminal distinction, the variants are presumably allophones of the Mandarin post-alveolar retroflex categories. According to the PAM, non-native speakers are expected to distinguish the retroflex variants from the alveolar counterparts very well when their languages have post-alveolar and alveolar

phonemes that involve the same phonation process (i.e., voiceless) and degree of constriction (i.e., fricative or affricate). That is, these speakers are likely to show TC assimilation for both types when the retroflex sounds are presented along with their alveolar counterparts. This prediction, however, is only preliminary. To generate more accurate predictions regarding non-native speakers' perceptions, it is necessary to examine the sounds closely related to the Mandarin contrasts in their native languages and the phonological functions of the sounds. The next section will explore the sound inventories of Japanese, English, and French. Figure 5 presents a number of the relevant sounds that all belong to the post-alveolar categories of the three languages. These sounds will be discussed in more detail.

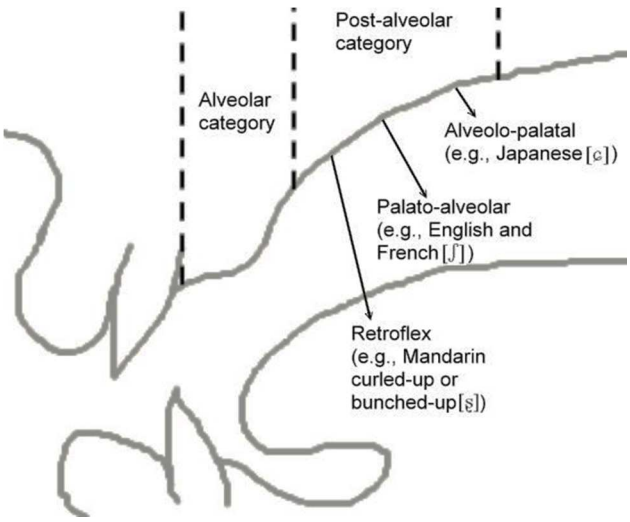

Figure 5 Mandarin, Japanese, English, and French sounds presented with their places of articulation in the post-alveolar region

2.3 Predictions for native speakers of Japanese, English, and French

2.3.1 Japanese

Spoken Japanese consists of segments in a relationship that are similar to Mandarin alveolars and retroflexes, and their phonological statuses may determine how Mandarin alveolars are distinguished from retroflexes, which include the CU and BU variants as their allophones. Akamatsu (1997) describes Japanese voiceless fricative [s] and affricate [ts] as lamino-alveolar (i.e., produced by raising the tongue blade towards the alveolar ridge) and fricative [ɕ] and affricate [cɕ] as laminodorso-alveolo-palatal (i.e., produced with the blade and front part of the tongue approaching the region between the alveolar ridge and hard palate). According to Best's (1995) description of the universal phonetic domain, the locations of constrictions of [ɕ] and [cɕ] are within the post-alveolar region and the sounds are likely to be regarded as post-alveolars together with the Mandarin [ʈʂ], [ʈʂʰ] and [ʂ].[5] Therefore, it can be inferred that if Japanese contrasts /s/ and /ts/ with /ɕ/ and /cɕ/, Mandarin alveolars and retroflexes—CU or BU—would be recognized as exemplars of the Japanese /s/ or /ts/ and /ɕ/ or /cɕ/, respectively. This recognition should result in the TC assimilation pattern on the part of Japanese listeners.

Such a prediction, nevertheless, remains open to debate as there is no general agreement regarding the relationship of the Japanese sounds. Akamatsu

[5] Best (1995) does not specify the locations of contraction specifically for retroflexes, palato-alveolars (e.g., [ʃ]), and alveolo-palatals; instead, she simply labels the region between the alveolar ridge and hard palate as post-alveolar, which suggests that these sounds should be collectively perceived as post-alveolars.

(2000) enumerates 29 Japanese phonemes, of which /s/, /ts/, /ɕ/, and /cɕ/ are single phonemes with their respective realizations as [s], [ts], [ɕ], and [cɕ]. However, Labrune (2012) considers [ɕ] as merely an allophone of /s/ given that it occurs when /s/ precedes /i/ or /j/. These divergent views lead to two different predictions. If Akamatsu's view were adopted, then the PAM would generate the aforementioned prediction that the CU and BU retroflexes are equally categorized into Japanese post-alveolar phonemes (such that there is no further phonological distinction between the apical and laminal sounds). If the post-alveolars were sub-phonemic to Japanese speakers, both retroflex variants should be categorized together as Japanese alveolar phonemes along with Mandarin alveolars. Then, the speakers are predicted to show Single Category (SC) assimilation or Category Goodness (CG) assimilation, such that thein which Mandarin alveolars should be better exemplars of the Japanese alveolar phonemes.[6] It is merely a matter of the category to which Mandarin retroflexes should be assimilated, as both retroflex variants are expected to be discerned equally well or poorly from their alveolar counterparts in either prediction. However, given that the unresolved issue regarding the phonological relationship of the Japanese sounds falls within the province of Japanese phonology and lies beyond the scope of the current study, both predictions are taken into consideration.

[6] Two non-native phones show SC assimilation when they are assimilated equally well into a native category. They show CG assimilation when one is perceived as a better exemplar of the category than the other (Best, 1995).

2.3.2 English

A correspondence with the Mandarin contrast in English, i.e., in the [s] and [ʃ] pair. Both are voiceless fricatives. The [s] sound is produced with the tongue tip or (along) with the blade (Malmberg, 1963; Kreidler, 1989; Ladefoged & Maddieson, 1996; Carr, 1999). With regard to [ʃ], some authors (e.g., Carr, 1999) regard it as a laminal, whereas others (e.g., Jones, 1972) contend that it can be either apical or laminal. The primary difference between the two English sounds is that the former is an alveolar, whereas the latter is a palato-alveolar articulated with the lips rounded or protruded. In addition to [ʃ], the affricate [tʃ] is also a palato-alveolar. Nevertheless, its alveolar counterpart [ts] is not a well-formed sequence in a syllable-initial position nor does /ts/ occur as an English phoneme. Only the sounds and phonemes that are native to the English language (i.e., [s] and [ʃ]) are addressed in the present discussion.

Unlike that of the Japanese sounds, the phonological relationship between the English /s/ and /ʃ/ seems unambiguous: /s/ and /ʃ/ are phonemes. Because the English /ʃ/ is also likely to be classified as post-alveolar according to the Best's (1995) characterization of the universal phonetic domain, English speakers would distinguish the two types of retroflexes equally well from the Mandarin alveolar sounds in the TC pattern, as they do for the English [s] and [ʃ].

2.3.3 French

The same TC assimilation is predicted for French speakers, who have alveo-dental vs. palate-alveolar contrasts. As with English, French draws a

phonemic distinction between /s/ and /ʃ/. However, their articulations in French are somewhat different from those in English. French [s] is an alveo-dental articulated with the tongue blade, and its [ʃ] is produced by rounding the lips and raising the tongue dorsum upward to form constriction between the post-alveolar and palatal regions (Fagyal, Kibbee, & Jenkins, 2006). Moreover, according to Fagyal *et al*., the affricates /tʃ/ and /ts/ are sounds in this contrast. Consequently, the Mandarin alveolars /ts/, /tsʰ/, and /s/ should correspond most closely to the French alveo-dental phonemes, and the Mandarin retroflexes /ʈʂ/, /ʈʂʰ/, and /ʂ/ should correspond most closely to the French palato-alveolar phonemes. French speakers are also expected to exhibit the TC type and discern the Mandarin alveolars from the retroflexes produced with the CU and BU gestures.

In sum, the discussion presented in this section suggests that non-native listeners who have distinct categories for alveolars and post-alveolars should be able to distinguish between Mandarin alveolars and retroflexes. Nevertheless, according to the PAM, it is not yet clear whether one retroflex variant promotes the distinction better than the other, despite their articulatory and acoustic differences. For this reason, an experiment testing the previously discussed PAM prediction is conducted.

3. Method

3.1 Experiment

Participants completed an ABX discrimination task in which they listened to three stimuli (A, B, and X) and determined whether X was equal to A or B.

Before the task was conducted, the sound tokens that were used in the study were recorded. Three Taiwanese Mandarin speakers who were capable of using both the CU and BU configurations and who had training in phonetics participated as sound informants. These Mandarin speakers provided sound tokens for the following six words by reading each word into a microphone in an anechoic room:

Table 1 List of six Mandarin words read by the informants

Initial consonant			Carrier vowel	Words (in IPA)
Retroflex	/ʈʂ/	+	/a/	/ʈʂa/$_{55}$
	/ʈʂʰ/			/ʈʂʰa/$_{55}$
	/ʂ/			/ʂa/$_{55}$
Alveolar	/ts/			/tsa/$_{55}$
	/tsʰ/			/tsʰa/$_{55}$
	/s/			/sa/$_{55}$

All of the words were consonant-vowel (CV) sequences with the same tone (i.e., tone 1) and with /a/ as the carrier vowel. The rationale for using the single vowel /a/ when constructing the words was that other vowels may contain rounded vowels (e.g., /ou/), which may affect the preceding consonant when coarticulation occurred or may have violated the phonotactic constraints of Mandarin (e.g., */ʈʂ/+/ei/). A number of other vowel contexts were less suitable than /a/ because they were not present in some of the languages (e.g., /aŋ/ does not occur in Japanese). To minimize the non-native speakers' bias favoring phonotactic structures more familiar to them, the common vowel /a/ was selected. Then, all of

the words were shown to the informants in Pinyin and Zhuyin. For the retroflex words (i.e., /ʈʂa/, /ʈʂʰa/, and /ʂa/), the informants were asked to articulate the words first with the CU and then with the BU gestures. This resulted in nine token types that could be divided into the following three groups: 1) CU /ʈʂa/, /ʈʂʰa/, and /ʂa/; 2) BU /ʈʂa/, /ʈʂʰa/, and /ʂa/; and 3) alveolar /tsa/, /tsʰa/, and /sa/. For each type, the informants produced five or six tokens, two of which were selected for use in the task.

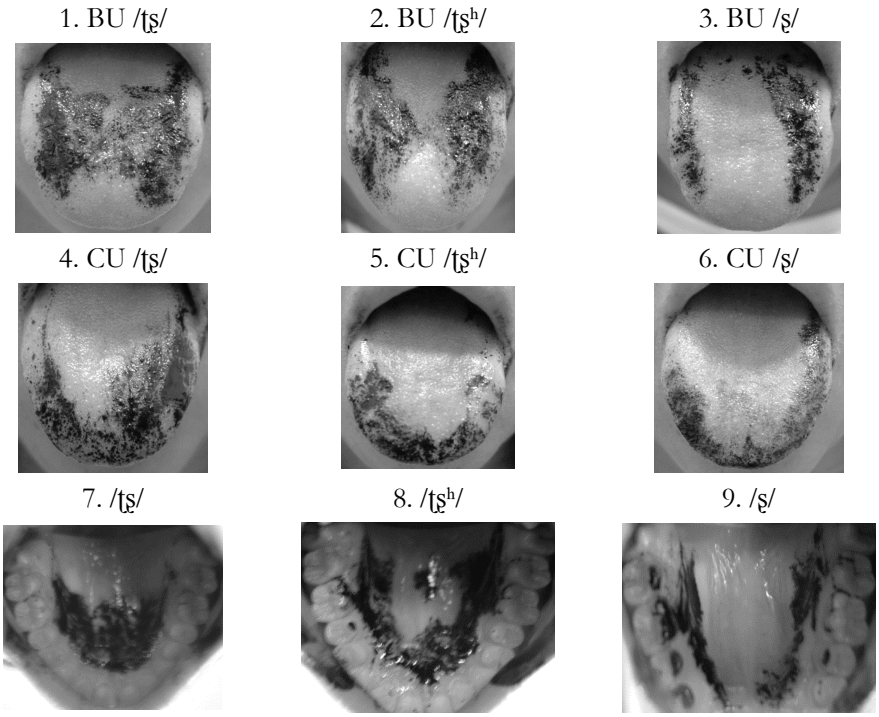

Figure 6 Linguograms and palatograms of the retroflexes produced in the bunched-up (BU) or curled-up (CU) tongue gesture

Palatography/linguography was conducted to examine whether the retroflex /ʈʂ/, /ʈʂʰ/, and /ʂ/ were produced in two different gestures, as required.[7] Charcoal powder mixed with olive oil was painted onto the palate and tongue of the speaker for recording the lingual contact patterns and palatographic patterns, respectively. The photos were taken using a digital camera. Figure 6 presents examples of selected linguograms and palatograms. The linguograms show that the blade of the tongue, but not the tip of the tongue, was raised in the BU type articulations (from 1 to 3), whereas both the tip and tiny part of the underside of the tongue, but not the blade of tongue, were involved in constriction in the CU type articulations (from 4 to 6). Since the palatographic data did not show significant differences between the two gestures, only three photos were presented to show that the post-alveolar region was contacted (from 7 to 9). These results suggest that the sounds produced using the BU gestures are laminal retroflexes, whereas the CU gestures are apical retroflexes.

The task was designed to present the tokens in the ABX format. A was always a token from the alveolar group, B was its retroflex counterpart from either of the retroflex token groups, and X was a different token of the same type as A or either a CU or BU retroflex counterpart of A. When a trial contained two stimuli of the same type (e.g., A = alveolar [tsa]; B = CU [ʈʂa]; and X = CU [ʈʂa]), two different tokens of that type were used. To avoid positional bias on the part of the participants, trials with their stimulus triads in BAX order were also included. In addition, filler trials were included. The filler stimuli were intended to alter the position of the target phone or the syllable structure of the experimental items

[7] We appreciate the reviewer's suggestion to conduct palatography/linguography to ensure that different gestures were used in the production of the post-alveolars.

(Mirman, McClelland, & Holt, 2006). the experimental items were all monosyllabic CV words with targets as their onsets, the filler trials consisted of two distinct phones of other syllable structures or target phones in other positions (e.g., first stimulus = [a]; second stimulus = [i]; and third stimulus = [i]). All three stimuli in a trial were produced by the same informant, with an interstimulus interval (ISI) of 1,000 milliseconds (ms). In total, there were 113 trials: Five filler trials plus 108 trials[8] that involved Mandarin alveolar vs. retroflex contrasts. The trials were divided into five blocks and presented to the participants in a random order.

3.2 Procedure

Participants were seated in front of a desktop computer and a response box in a quiet room. They were asked to wear a pair of headphones. An audio file for volume testing was played until the participants were comfortable with the volume. At the outset of the experiment, their language backgrounds and experiences with learning Chinese were recorded. Then, they were instructed regarding how to respond to the stimuli in each trial. They were asked to judge which word was the same as the third word by pressing the associated key: Key "1" should be pressed if they considered the first and third words to be the same word and key "2" should be pressed if they considered the second and third words to be the same word.

---

[8] Six basic permutations of ABX × three consonant types (unaspirated affricate, aspirated affricate, and fricative) × three informants × two orders (ABX or BAX)

3.3 Participants

The participants were 10 native Taiwan Mandarin speakers and 42 non-native speakers of Mandarin who were learning elementary Chinese in beginners' classes at a university in southern Taiwan. The native speakers were assigned to the control group and the non-native speakers were assigned to the experimental group. The experimental group was further divided into groups according to the participants' L1 backgrounds, and the three major groups were Japanese, English, and French (which accounted for 35 of the participants). The three groups could not be balanced equally for number. The numbers of participants recruited were 14 for the Japanese, 10 for the English, and 11 for the French groups. The mean number of months spent learning Chinese according to self-report was 26.7 for the Japanese, 20.3 for the English, and 12.8 for the French groups. The remaining seven participants were two Korean speakers and five speakers of Singapore English.

4. Results

Statistical analyses were performed on the data obtained from the Taiwan Mandarin, Japanese, English, and French participants.[9] One datum of the Japanese group was deleted because the participant failed to respond correctly to

[9] Due to an insufficient number of participants, the data for the Korean and Singapore English groups may not be appropriate for statistical analyses.

all five filler trials. The remaining data were then analyzed different patterns in which a “same” response could have been made to any two tokens in a trial. A same response was correct when it was made to 1) a BU retroflex token and a CU retroflex token, 2) two BU retroflex tokens, 3) two CU retroflex tokens, and 4) two alveolar tokens. It should be noted that the first response pattern was correct because the CU and BU words were expected be distinguished from the alveolar (ALV) word in the same trial. These four response patterns are abbreviated as BU-CU, BU-BU, CU-CU, and ALV-ALV. For each pattern, the percentage of correct responses, or the accuracy rate, was calculated by dividing the number of participants’ responses made in a specific pattern by the total number of trials in which that pattern was possible.

Table 2 presents the accuracy rates for the BU-CU response type for the four language groups. To determine whether the BU and CU retroflexes were classified together and distinguished from the alveolars equally well by the four groups, a one-way ANOVA was conducted to compare the percentages of these response types. The analysis revealed a significant difference, $F(3, 40) = 5.12$, $p < .01$. Post-ANOVA pair-wise comparisons using Tukey HSD tests indicated that the average accuracy rate for the Japanese group was significantly lower than those for the English ($p < .05$) and French ($p < .01$) groups.

To further investigate whether either gestural variant may be better distinguished from their alveolar counterparts by speakers within the same group, several paired *t*-tests were conducted. The tests compared the mean accuracy rates for the BU-BU and CU-CU responses (shown in Table 3) and yielded no significant differences in accuracy rates between the two response types for the Japanese ($t(12) = 1.31$, $p > .10$, two-tailed), English ($t(9) = -0.81$, $p > .10$,

two-tailed), and French ($t(10) = 0.64$, $p > .10$, two-tailed) groups. However, a significant difference was evident for the Taiwan Mandarin group ($t(9) = -4.71$, $p < .01$, two-tailed).

Finally, the consonant onsets of the 18 tokens provided by the informants were acoustically analyzed in terms of M1. The results indicated that the mean M1 value was 2.91 kHz for the CU tokens (SD = 0.60 kHz), 3.56 kHz for the BU tokens (SD = 0.40 kHz), and 6.09 kHz for the alveolar tokens (SD = 1.58 kHz). Thus, the BU and CU tokens were acoustically different.

Table 2 The average accuracy rates and standard deviations (SD) for the BU-CU response type for the four L1 groups

L1 groups	*n*	BU-CU pattern	
		Mean accuracy rate	SD
Taiwan Mandarin (control)	10	83.89%	13.34
Japanese	13	71.58%	13.38
English	10	86.52%	12.41
French	11	89.14%	7.5

Table 3 The average accuracy rates and standard deviations for the BU-BU and CU-CU response patterns for the four groups.

	n	BU-BU pattern		CU-CU pattern	
		Mean accuracy rate	SD	Mean accuracy rate	SD
Taiwan Mandarin (control)	10	83.89%	12.68	92.78%	9.09
Japanese	13	77.78%	13.22	74.16%	21.06
English	10	88.33%	9.60	90.00%	8.20
French	11	92.42%	5.14	91.38%	6.25

5. Discussion

5.1 Interpretations of the results

The results obtained from the ABX discrimination task appear to be consistent with the PAM predictions for the English and French participants. The mean accuracy rates for the BU-CU response type for the English and French groups suggest that both groups generally perceived the BU and CU retroflexes as from the same category, which differed from the category in which the Mandarin alveolars were classified. Within each of these groups, the two retroflex variants were also equally distinguished from the alveolars because no significant difference was obtained ($p > .10$). These findings provide evidence that the English and French participants show TC assimilation for the Mandarin alveolar vs. retroflex contrasts regardless of the gestural configuration of the retroflex. Therefore, for English and French listeners, the CU-BU variation may not interfere with the phonological perception of the Mandarin contrasts in question.

Before the Japanese participants' data are interpreted, the results for the control group reveal two curious findings regarding the Taiwan Mandarin speakers' perception of the CU and BU retroflexes. First, despite being native speakers, the Taiwan Mandarin participants did not reach perfect or near perfect accuracy rates when making BU-CU responses. Second, unlike participants from the other L1s, the Taiwan Mandarin participants were generally more accurate when responding to the CU-CU response pattern (accuracy rate = 92.78%) compared to the BU-BU one (accuracy rate = 83.89%). One possible explanation

for these observations is that the native speakers of Taiwan Mandarin are able to perceive the articulatory-acoustic differences between the retroflex variants, which in turn affect their perception of the alveolar vs. retroflex contrasts. An interpretation of the results under this assumption suggests that the BU variant is the comparatively poorer exemplar of the retroflex category. This interpretation can account for the Taiwan Mandarin speakers' failure to achieve perfect accuracy when responding to the BU-CU trials: The poorer goodness-of-fit of the BU retroflex tokens occasionally led them to generate erroneous responses. In addition, the assumption that the CU retroflex variant is perceptually more retroflexed for native Taiwan Mandarin listeners is consistent with the results of the acoustic measurement, which suggest that the CU retroflexes have lower M1 values than the BU ones.

Indeed, there is indirect evidence that the acoustic properties of the BU retroflexes may support its poorer distinction from the alveolars by native speakers of Taiwan Mandarin. Cheng (2009) investigates Taiwan Mandarin speakers' discrimination of three categories of monosyllabic words that vary in the degree of retroflexion of the consonant onset and finds that 61.1% of the "slightly retroflexed" stimuli are perceived as retroflex, whereas the other 39.9% are perceived as non-retroflex (alveolar). The slightly retroflexed stimuli belong to an in-between category given that its M1 value is higher than that of the "more retroflexed" category but lower than that of the "non-retroflex" category. The slightly retroflexed stimuli in Cheng's study are similar to the BU retroflex tokens utilized in the current study, which have an average M1 value that is between the values for the CU tokens and alveolar tokens.

The Japanese participants exhibit a slight tendency towards TC assimilation,

although the data appear to undermine this claim when the accuracy rates for the four groups are compared. The Japanese group's percentage of correct responses for the BU-CU response pattern is significantly lower than that of the other two groups ($p < .01$); nevertheless, this finding cannot be attributed to either retroflex variant given that the CU and BU retroflexes are not significantly different ($p > .10$). The question then arises as to whether accuracy rates of approximately 70% can be used as evidence for any PAM prediction. Currently, determining whether the accuracy rates are high enough when compared to those of the control group is a tricky issue that is complicated by the possibility that the Taiwan Mandarin speakers have a preference for CU retroflexes compared to BU ones, which is a phenomenon that is not observed in any of the non-native groups. Clues regarding the articulatory-acoustic differences between the retroflex variants may influence the Taiwan Mandarin participants' accuracy when responding to the BU-CU, BU-BU, and CU-CU response patterns. Thus, the average accuracy rates for the control group may not be directly comparable with those for any of the other non-native groups. Whether a point of reference for accuracy is available or not, if the average percentages of correct responses for the Japanese group are considered sufficiently high, it is possible that the Japanese participants assimilated the Mandarin alveolars and retroflexes into two different native categories and that Akamatsu's (2000) characterization of /ɕ/ and /cɕ/ as single phonemes appropriately describes the phonological statuses of the two sounds.

If this view were accepted, it is necessary to account for the gap in accuracy rates between the Japanese group and the other non-native participants. One possible explanation provides an articulatory-phonetic account for this

phenomenon: The Japanese participants may perceive greater phonetic deviations from their "post-alveolar" phonemes in the Mandarin retroflexes than the English and French participants. As previously illustrated, Japanese alveolo-palatals and English and French palato-alveolars are given the blanket label "post-alveolar" when the PAM predictions are proposed. However, the sounds differ in their exact articulation. For example, unlike the palato-alveolar [ʃ], the alveolo-palatal [ɕ] is palatalized (Ladefoged & Maddieson, 1996). It may be more difficult for the Japanese participants to treat the retroflexes as good exemplars of their alveolo-palatal categories after they notice the absence of palatalization in the Mandarin sounds. In contrast, the English and French [ʃ] and [tʃ] are only weakly palatalized; therefore, native speakers of these languages may not consider the retroflexes as deviating from /ʃ/ and /tʃ/, which may explain why the accuracy of the Japanese group is poorer compared with that of the other two groups. This explanation relies on the presumption that phonetic properties, such as palatalization, may affect the categorical perception of non-native phones.

Given the assumption that phonetic properties may interfere with the categorization of non-native sounds, it is tempting to conclude that the non-native speakers are unable to distinguish between the CU and BU retroflexes because their BU-BU and CU-CU response patterns are not different. However, it is necessary to step back and first consider the experimental design of the ABX discrimination task and the extent of the differences. Strange and Shafer (2008) discuss a study by Werker and Logan (1985) and indicate that listeners respond to non-native phones based on categorization into native phonemes in highly demanding tasks (e.g., when the interstimulus interval is long). On the other hand, in tasks that require minimal cognitive effort, listeners are able to discern acoustic

differences between two tokens that do not differ phonologically. In other words, the experimental design may have determined whether listeners relied more on phonetic or phonological modes when processing upcoming stimuli. If the task in the current study was less demanding, for example, by shortening the interstimulus interval, participants may notice the phonetic differences between the CU and BU retroflexes. This variable is experimental, which, as Strange and Shafer note, is not addressed in the PAM. A consideration of this factor undermines the previous conclusion regarding the CU and BU retroflexes since the data do not provide concrete evidence that the non-native participants are unable to perceive at least a modicum of the phonetic difference between the two retroflex variants. Nevertheless, one conclusion can be drawn from the discussion thus far: Regarding the phonological perception of the Mandarin alveolar vs. retroflex contrasts, the two types are not different for non-native speakers.

5.2 Implications for Mandarin pedagogy

The previously discussed interpretations serve as the basis for the pedagogical suggestions proposed here to help non-native speakers learn the Mandarin retroflexes. Therefore, it is important to first specify what type of pedagogical issues these suggestions address. One issue is intelligibility, which is addressed in Jenkins' (2000) concept of EIL. Jenkins argues for an emphasis on intelligibility among non-native speakers of English, such that instructors can treat Mandarin Chinese as an international language and prioritize enhancing intelligibility among non-native learners.

If the focus of Mandarin teaching and learning was on intelligibility, the

previous discussion of the results highlights the need to assist Japanese learners' discrimination of Mandarin alveolars and retroflexes. One way to improve this discrimination is to utilize a heuristic approach in which students are constantly exposed to the phonetic features of Mandarin retroflexes. One assumption is that the lower accuracy rates for the Japanese participants can be attributed to less goodness-of-fit of the retroflexes to Japanese alveolo-palatal phonemes, causing the Japanese participants to ignore the lack of palatalization. This assumption is interpreted to mean that Japanese speakers' attention to phonetic differences between native and non-native phones could be reinforced to help them recognize the Mandarin retroflexes as sounds from new categories. Moreover, the formation of these new categories is not without a theoretical basis. According to the Speech Learning Model (SLM) proposed by Flege and his colleagues (e.g., Flege, 1995), a new phonetic category can be established for a non-native sound when L2 learners perceive its difference from the closest native phone. As a result, Mandarin instructors could encourage their Japanese learners to discover the phonetic properties of the retroflexes by frequently comparing these properties with those of alveolars and alveolo-palatals (e.g., /ɕ/, a sound that occurs in both Mandarin and Japanese). Japanese learners should show improvement in their perception of the Mandarin contrasts after a retroflex category is established that is based on the perceived differences between the alveolar and alveolo-palatal categories.

Once the problem of intelligibility is solved, instructors can focus on the articulatory aspects of Mandarin retroflex pronunciation and determine which retroflex variant, either the CU or BU, requires less articulatory effort for non-native learners with a particular L1 background. For these two types of retroflexes, the degree of articulatory difficulty may be predicted by their gestural

proximity to certain sounds in a learners' native language. For example, as described in studies examining the phonetics of English (e.g., Ladefoged & Maddieson, 1996) and French (e.g., Fagyal, *et al*., 2006), the tongue shape for the [ʃ] sound is "domed."[10] If English or French speakers produce [ʃ] without lip-rounding and find that it sounds similar to the Mandarin retroflex [ʂ], it could be that the BU gesture is easier because it is also articulated by bunching up the tongue body. By contrast, speakers of other languages may consider the CU gesture to be easier. For example, according to Ladefoged and Maddieson (1996), sub-apical retroflexes, such as [ɖ] and [ʂ], involve a considerable amount of tongue-tip curling and occur in a number of Dravidian languages (e.g., Tamil, Telugu, and Toda). L2-Mandarin learners of these languages may transfer their native retroflex articulation to the L2 and produce the CU retroflexes effortlessly. Currently, these hypotheses regarding articulatory ease are simply inferences based on studies of phonetics. These hypotheses need to be confirmed by further investigations or supported by testimony from Mandarin instructors or learners. However, Mandarin instructors should consider both retroflex types and allow their students to opt for the easier type while adhering to most of the pedagogical materials and teaching only the CU retroflexes.

5.3 Research limitations

Interpretations of the results and pedagogical suggestions are appropriate only to a certain extent because they are subject to specific variables that must be acknowledged and controlled for in future research. The sources of these

[10] English [ʃ] is domed when it is laminal.

variables are the participants and the experimental paradigm. First, the three groups should be balanced and the number of participants in each group should be increased such that parametric statistical analyses (e.g., ANOVA) are appropriate for the data. Additionally, extraneous factors, such as learning experience, should be excluded. Information about the participants[11] that was recorded at the beginning of the experiment reveals that they differed slightly in their experiences when learning Mandarin. This factor may contribute to differences among the groups and may even provide an alternative explanation for the lower accuracy rates for the Japanese participants, i.e., their performance was poorer due to their inexperience. Finally, as previously discussed, a listening experiment may elicit responses that are based on phonological categorization or phonetic perception, depending on the design. The ABX task utilized in the current study induced categorical perception of the stimuli and did not provide evidence that the non-native participants were insensitive to the phonetic differences between the CU and BU retroflexes. This variable and the participant-related factors should be considered in future research.

6. Conclusion

The present study investigated whether articulatory differences between the CU and BU retroflexes led to perceptual differences in the Mandarin alveolar vs. retroflex contrasts for non-native speakers. The results revealed that the

[11] Unfortunately, the exact ages and the ages of learning of the participants were not recorded during the experiment; however, their estimated ages range between 20- and 30-years-old.

non-native participants generally showed Two Category assimilations for the contrasts regardless of the retroflex gestures. However, the accuracy rates for the Japanese group were comparatively lower than for the other two groups, which might be due to their attention to the phonetic property of palatalization. Based on the interpretations of the results, a number of suggestions were proposed to promote intelligibility among Japanese speakers and to facilitate non-native speakers' learning of retroflex pronunciation in general.

References

Akamatsu, Tsutomu, *Japanese Phonetics: Theory and Practice* (日語語音學之理論與實踐), München, Newcastle: Lincom Europa, 1997.

Akamatsu, Tsutomu, *Japanese Phonology: A Functional Approach* (以功能研究法探討日語音韻學), München, Newcastle: Lincom Europa, 2000.

Ball, Martin J. and Joan Rahilly, *Phonetics: The Science of Speech* (語音學：言語的科學), London: Arnold, 1999.

Best, Catherine T., Gerald W. McRoberts, and Elizabeth Goodell, Discrimination of Non-native Consonant Contrasts Varying in Perceptual Assimilation to the Listener's Native Phonological System (非母語聽者對同化到其母語音韻系統的子音對立的區辨), *The Journal of the Acoustical Society of America*, *109*, 2001, 775-794.

Bickford, Anita C. and Rick Floyd, *Articulatory Phonetics: Tools for Analyzing the World's Languages* (發音語音學：分析世界語言的工具), Dallas, Texas: SIL International, 2006.

Carr, Philip, *English Phonetics and Phonology: An Introduction* (英語語音學與音韻學之導論), Malden Massachusetts: Blackwell Pub, 1999.

Delattre, Pierre, The Physiological Interpretation of Sound Spectrograms (生理學上對聲譜圖的詮釋), *Publications of the Modern Language Association of America*, *65*, 1951, 864-875.

Fagyal, Zsuzsanna, Douglas Kibbee, and Frederic Jenkins, *French: A Linguistic Introduction* (法語語言學導論), Cambridge: Cambridge University Press, 2006.

Flege, James Emil, Second-language Speech Learning: Theory, Findings and

Problems (二語學習之理論、發現與問題), *Speech Perception and Linguistic Experience: Issues in Cross-linguistic Research*, Timonium: York Press, 1995, 229-273.

Guenther, Frank H., Carol Y. Espy-Wilson, Suzanne E. Boyce, Melanie L. Matthies, Majid Zandipour, and Joseph S. Perkell, Articulatory Tradeoffs Reduce Acoustic Variability during American English /r/ Production (構音時的權衡取捨減少美式英語/r/發音在聲學上的變異性), *The Journal of the Acoustical Society of America*, *105*, 1999, 2854-2865.

Guo, Zhe-chen and Shu-chen Ou, Perception of Articulatorily Different Mandarin Retroflexes by Japanese Speakers: A Pilot Study (日語母語人士對構音方式不同的中文捲舌音的感知：初步研究), paper presented at The 2nd NTU Postgraduate Conference on Teaching Chinese as a Second Language, Taipei, 2014.

Jakobson, Roman, C. Gunnar M. Fant and Morris Halle, *Preliminaries to Speech Analysis: The Distinctive Features and their Correlates* (初步語音分析：辨音成分與語音關聯), Cambridge, Mass.: Massachusetts Institute of Technology Press, 1951.

Jenkins, Jennifer, A Sociolinguistically Based, Empirically Researched Pronunciation Syllabus for English as an International Language (以社會語言學為基礎並經實證研究，為英語做為國際語言所提出的發音大綱), *Applied Linguistics*, *23*, 2002, 83-103.

Jones, Daniel, *An Outline of English Phonetics, 9th Edition* (英語語音學概述), New York: Cambridge University Press, 1972.

Kreidler, Charles W., *The Pronunciation of English: A Course Book in Phonology* (英語的發音：音韻學教程), Oxford: B. Blackwell, 1989.

Labrune, Laurence, *The Phonology of Japanese* (日語音韻學), New York: Oxford University Press, 2012.

Ladefoged, Peter and Ian Maddieson, *The Sounds of the World's Languages* (世界語言的聲音), Oxford: Blackwell Publishers, 1996.

Ladefoged, Peter and Keith Johnson, *A Course in Phonetics, 6th Edition* (語音學), Boston: Thomson Wadsworth, 2011.

Ladefoged, Peter and Zong-ji Wu, Places of Articulation: An Investigation of Pekingese Fricatives and Affricates (構音部位：北京話中擦音與塞擦音之研究), *Journal of Phonetics, 12*, 1984, 267-278.

Ladefoged, Peter, *A Course in Phonetics, 5th Edition* (語音學), Fort Worth, Texas: Harcourt College Publishers, 2001.

Lee, Wai-Sum and Eric Zee, Standard Chinese (Beijing) (標準中文（北京話）). *Journal of the International Phonetic Association, 33*, 2003, 109-112.

Lin, Yen-hwei, *The Sounds of Chinese* (華語的聲音), New York: Cambridge University Press, 2007.

Lindau, Mona, Vowel Features (母音的特徵), *Language, 54*, 1978, 541-563.

Malmberg, Bertil, *Phonetics* (語音學), New York: Dover Publications, 1963.

Mirman, Daniel, James L. McClelland, and Lori L. Holt, Attentional Modulation of Lexical Effects on Speech Perception: Computational and Behavioral Experiments (詞彙對語言辨識之影響的注意力調控：計算與行為實驗), Proceedings of the 28th Annual Cognitive Science Society, New Jersey, 2006.

Strange, Winifred and Valerie L. Shafer, Speech Perception in Second Language Learners: The Re-education of Selective Perception (二語學習者的語言感知：選擇性知覺的再教育), *Phonology and Second Language Acquisition.* Philadelphia: John Benjamins, 2008, 153-191.

Wan, I-Ping and Jeri J. Jaeger, The Phonological Representation of Taiwan Mandarin Vowels: A Psycholinguistic Study (從心理語言學的角度研究臺灣華語中母音的聲韻表徵), *Journal of East Asian Linguistics*, *12*, 2003, 205-257.

Werker, Janet F. and John S. Logan, Cross-language Evidence for Three Factors in Speech Perception (語言感知的三要素：跨語言的證據), *Perception & Psychophysics*, *3*, 1985, 35-44.

Zee, Eric, The Phonetic Characteristics of the Sounds in Standard Chinese (Beijing) (標準中文（北京話）中聲音的語音特性), paper presented at the 15th North American Conference on Chinese Linguistics, Lansing, 2003.

Zhou, Xinhui, Carol Y. Espy-Wilson, Suzanne Boyce, Mark Tiede, Christy Holland, and Ann Choe, A Magnetic Resonance Imaging-based Articulatory and Acoustic Study of 'Retroflex' and 'Bunched' American English /r/. *The Journal of the Acoustical Society of America*, *123*, 2008, 4466-4481.

王若江：《漢語正音教程》，北京：北京大學，2005 年。

李曉琪、戴桂芙、郭振華：《新漢語教程：情景、功能、結構》，北京：北京大學，1995 年。

國立臺灣師範大學國音教材編輯委員會編：《國音學》，臺北：正中，2007 年。

陸慶和：《對外漢語教學實用語法》，北京：北京大學，2006 年。

葉德明：《初級華語十課》，臺北：世界華文教育協進會，1994 年。

鄭靜宜：〈國語捲舌音和非捲舌音的聲學特性〉，《南大人文研究學報》第 40 期（2006 年），頁 27-48。

鄭靜宜：〈華語捲舌音對比的聽覺辨識與頻譜動差分析〉，《中華心理學刊》第 2 期（2009 年），頁 157-174。