Dialectal variation in the perception of phonological contrasts

Yung-hsiang Shawn Chang[a)]
Department of English, National Taipei University of Technology, Taiwan

Chilin Shih
Department of Linguistics, University of Illinois at Urbana-Champaign, IL 61822

Jont B. Allen
Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, IL 61801

(Dated: November 6, 2013)

The present study examined the cross-dialectal variability in perceptual patterns of the alveolar-retroflex contrast, a critical feature often used to differentiate Beijing Mandarin from other dialects of Mandarin, such as Taiwan Mandarin. While previous research has focused on dialectally and contextually driven variation in alveolar and retroflex productions, little is known about whether similar variability is also present in perception. We investigated the categorical and gradient modes of alveolar-retroflex perceptions in different vowel contexts by listeners from Beijing and Taiwan. The results indicate that that Beijing and Taiwan listeners have different perceptual boundaries along the acoustic continuum, with a lower cutoff frication frequency required for the retroflex percepts for Beijing listeners. Both groups of listeners' category boundaries shift to lower frequencies in the rounded vowel context to normalize for vowel coarticulatory effects. Discrepant within-category sensitivity was found in that while both Beijing and Taiwan listeners perceived all retroflex variants as equally good exemplars, Beijing listeners considered the endpoint variant of the alveolar as the preferred category exemplar. We concluded an articulatory explanation for this asymmetrical within-category discriminability and also discussed the findings with respect to the hyperspace effect in perception. Together, our results show that linguistic (i.e., vowel context) and sociolinguistic (i.e., dialect) factors collectively and variably affect the perception of the Mandarin alveolar-retroflex contrast.

PACS numbers: 43.70.Mn, 43.71.Es

I. INTRODUCTION

While considerable work has been done on cross-language perception of non-native phonological contrasts, speech processing by listeners with a different dialect of the same language has only been recently investigated. Most available cross-dialect perception studies have primarily examined 1) how well listeners with and without contrast merger in their dialect can make perceptual discriminations (e.g., Janson and Schulman, 1983; Labov *et al.*, 1991), 2) what affects listeners' ability to identify dialectal pronunciations (e.g., Clopper and Pisoni, 2004; Flanigan and Norris, 2000; Preston, 1993; Sumner and Samuel, 2009), or 3) how social stereotyping can cause listeners to make perceptual adjustments (e.g., Evans and Iverson, 2004; Niedzielski, 1999). In studying the dialect effects on perception, an emerging line of research delves into the variability in sound categorization across dialect (Cutler *et al.*, 2005; Kendall and Fridland, 2012; Kraljic *et al.*, 2008). In the present study, we extended similar investigation of cross-dialect variability to the perception of the contrast between the alveolar (/s, ts, ts^h/) and retroflex (/ʂ, tʂ, $tʂ^h$/) consonants[1] in Mandarin. Previous research has reported patterns of variation in the production of Mandarin alveolar-retroflex contrast as a result of vowel context (Jeng, 2006; Li, 2009; Chang and Shih, 2012), prosodic prominence (Chuang and Fon, 2010), as well as sociolinguistic factors like dialectal background (Chang, 2011a; Chang and Shih, 2012) and speech style (Jeng, 2006). In view of such variation in production, this study aimed at investigating whether a dialect-specific pattern is present in perception and whether variation in perception can be attributed to both sociolinguistic and linguistic factors.

II. BACKGROUND

A. Mandarin alveolar-retroflex contrast: social indexing and cross-dialect variation

Standard Mandarin is based on the pronunciation of the Beijing dialect, and is the official language spoken

[a)] Author to whom correspondence should be addressed. Electronic mail: shawnchang@ntut.edu.tw

[1] Mandarin alveolar sibilants have been variously termed as dentals (Chao, 1968), alveolars (Kratochvil, 1968; Ladefoged and Maddieson, 1996; Luo and Wang, 1981), and denti-alveolars (Lee and Zee, 2003). On the other hand, Mandarin retroflex alveolars have been considered to be laminal post-alveolars (Ladefoged and Maddieson, 1996), apical post-alveolars (Lee and Zee, 2003) or just retroflexes (Duanmu, 2000). In this study, these two categories of sounds are consistently called alveolars and retroflexes.

in China and Taiwan. In Standard Mandarin, there exists a place contrast between the alveolar and retroflex sibilants. They both occur only in syllable-initial position, and are in contrastive distribution with each other in that both series can only be followed by /a, u, ə, i/ vowels and the [w] glide. It should be noted that the /i/ vowel surfaces as a high central vowel [ɪ] after the alveolar sibilants and as a high back vowel [ɹ] after the retroflex sibilants. There is a considerable vowel quality difference between [ɪ] and [ɹ] in terms of F2 and F3. However, phonologically, they are considered allophones of the same vowel.

The alveolar-retroflex contrast is a critical feature that distinguishes the standard Mandarin pronunciation from a local dialect-accented Mandarin pronunciation. The retroflex sibilants are absent in southern dialects[2] of Chinese, which has been argued to be the reason why southern Mandarin speakers' alveolar-retroflex production is subject to contrast neutralization (e.g., Chien, 1971 and Kubler, 1985 for Taiwan Mandarin; Zhu, 2012 for Shanghai Mandarin). On the other hand, the alveolar-retroflex contrast has been described to be more stable and consistent in Beijing Mandarin, a northern dialect of Mandarin (Duanmu, 2000; Lin, 2007). In the case of alveolar-retroflex neutralization, alveolars are the default forms. Therefore, the presence of retroflexion in speech can be used to determine whether one makes the place distinction.

In China and Taiwan, the use of retroflexion in speech is indexed for standard pronunciation and is associated with a higher education level. The alveolar-retroflex contrast is explicitly taught in school in China and Taiwan and has been promoted in various ways. In China, a standardized oral proficiency at the state level, Putonghua Shuiping Ceshi (National Common Speech Proficiency Test), is administered to assess the oral proficiency of Chinese nationals if they wish to apply for jobs in public domains (e.g., schools and government administration). One of the indicators for intermediate-level speakers is to make inconsistent or no alveolar-retroflex distinction. In Taiwan, retroflexion is prescribed as a feature of textbook Mandarin, and is the standard for most media broadcasting (Chung, 2006).

To study the dialect effects on perception of the alveolar-retroflex contrast, we chose Beijing Mandarin and Taiwan Mandarin because 1) these two dialects are the major regional varieties of Mandarin, and 2) they are respectively representative of northern and southern dialects of Mandarin. Between the two dialects, different degrees of retroflexion in production have been found (Chang, 2011a). Different magnitudes of the alveolar-retroflex contrast and sibilant realizations have also been reported (Chang and Shih, 2012). On the basis of a strong link suggested between sociophonetically driven differences found in production and perception (Brunelliére *et al.*, 2009; Evans and Iverson, 2007), we expect the presence of cross-dialect variability in the perception of the alveolar-retroflex contrast.

[2] The word "dialect" is used in two senses in this study: "Chinese dialects" refers to various dialects of the Chinese languages, which include Mandarin and non-Mandarin dialects. Mandarin "dialects" refer to various regional pronunciations of Mandarin.

B. Context-conditioned variation in perception: categorical and gradient

In a recent production study, Chang and Shih (2012) found the vowel context to interact with the dialectal background (Beijing Mandarin vs. Taiwan Mandarin) and variably affected the realization of the alveolar-retroflex contrast. Specifically, Beijing Mandarin speakers produced a greater alveolar-retroflex contrast than Taiwan Mandarin speakers in the /a/ context, and the /u/ context was where both dialects exhibited a lesser degree of the place distinction. In view of such vowel context effects on production patterns across dialect, we added the vowel context as another independent variable in the present perception study.

Regarding context-induced variation in perception, research has shown that listeners are remarkably attuned to vowel coarticulation and would compensate for that variation in consonant perception (e.g., Whalen, 1979; Mann and Repp, 1980). Much of the evidence for contextual effects in perception is drawn from studies examining the location of category boundaries. The general finding is that boundary locations are flexible; context-induced variation results in a systematic change in the boundary location in perception. Whalen (1979) had participants listen to fricative noise tokens synthesized to represent a /ʃ-s/ continuum followed by either /i/ or /u/ and asked subjects in a force-choice experiment to label them as "s" or "sh". His data indicated a lower /ʃ-s/ boundary (in Hz) for /u/ than for /i/. Mann and Repp (1980) replicated Whalen's study by using synthetic fricative noises from a /ʃ-s/ continuum followed by /a/ or /u/. They also found the phoneme boundary to shift toward lower noise frequencies in the /u/ context. In addition, the results of their identification task showed that listeners perceived more instances of /s/ in the context of /u/, especially in the case of a fricative noise ambiguous between /s/ and /ʃ/. With that, they suggest that listeners are able to perceptually compensate for these coarticulatory effects.

Besides being reflected in the change of categorical boundary location, contextual effects have also been shown to alter the internal structure of a category in perception (see Allen and Miller, 2001; McMurray *et al.*, 2002; Miller, 1994; Pisoni and Tash, 1974 for discussion on subphonemic variation and perceptual exemplars). Kawasaki *et al.* (1986) had their English-speaking subjects rate the nasality of English /ɪ, u, a/ vowels in the /m_m/ context. The amplitude of the nasal consonant was attenuated in five steps to create the nasal vs. oral (or less nasal, as the amplitude of /m/ was more attenuated) contexts. They found that listeners gave a higher nasality rating for the nasalized vowels that occurred in the oral context than in the nasal context. That is, the same nasalized vowels were perceived as more nasalized when they were flanked by weaker nasal consonants.

The graded internal category structure was also found in voicing series (specified by voice onset time) varying in speech rate (Miller and Volaitis, 1989; Volaitis and Miller, 1992; Wayland *et al.*, 1994). A change in target-syllable rate would alter the voiced-voiceless boundary location as well as the within-category structure. Specifically, when speech rate was slowed, higher goodness-rating scores would be given to /p/ with longer voice onset time values, and a wider range of stimuli were considered as good exemplars of /p/ (Volaitis and Miller, 1992).

Taken together, the above studies indicate that listeners are aware of various types of systematic and context-dependent variation and are able to adjust the perceptual threshold and judgments along a given acoustic dimension. A change in context would not only alter category boundary location but also within-category structure. In this regard, in investigating perceptual variation, we elicited perceptual judgments on both discrete and continuous scales.

C. Perception cues for coronal contrasts

In distinguishing the place of articulation of coronal fricatives, the location of acoustic energy in the spectrum is the most often noted acoustic cue (e.g., Harris, 1958; Heinz and Stevens, 1961; Hughes and Halle, 1956; Shadle, 1985). To take English coronal fricatives for example, most energy in the noise spectrum of /s/ is located above 4000 Hz. In /ʃ/, the spectral peak occurs at about 3.5 kHz. Figure 1 displays the AI-grams[3] of /sa/ and /ʃa/ produced by a female native speaker of American English. The abscissa of the AI-grams marks the acoustic event in time (in centisecond) and the ordinate marks frequency (kHz), on a critical band scale. The noise band from around 10 to 25 centiseconds in the AI-gram indicates the fricative. While the alveolar /s/ has high-frequency signals above 4kHz, this frication region (the framed part in the AI-gram of /ʃa/) is also present in /ʃa/. Li and colleagues (Li and Allen, 2011; Li *et al.*, 2012) had their subjects listen to the /ʃa/ tokens with the frication noise above 4kHz removed, yet strong percepts of /ʃa/ were still reported. On the other hand, when the frication region between 2-4 kHz was removed from /ʃa/ by a time-varying high-pass filter, it was perceived as identical to the /sa/ token produced by the same talker. Modifying the frequency of the lowest bound of the wideband noise while leaving the rest of the frication intact results in the perception of a different fricative. Li and colleagues therefore concluded that the necessary perceptual cue for fricative place of articulation lies in the frequency of the lowest bound of the wideband noise (which will be termed "the cutoff frequency" hereafter). Raising the cutoff frequency of /ʃ/ by removing the frication region between 2-4 kHz will convert /ʃ/ into natural-sounding /s/.

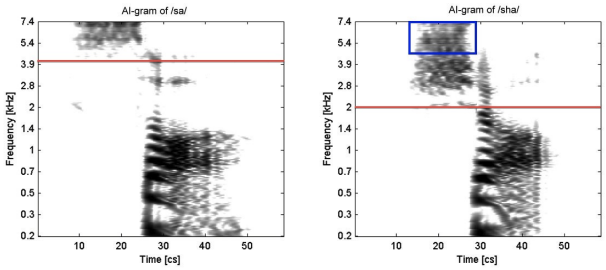


FIG. 1. AI-grams of /sa/ and /ʃa/ in English (adapted from Li and Allen (2011)). The time units are in centiseconds.

The most important acoustic attribute that distinguishes Mandarin /s/ and /ʂ/ lies in the frequency domain as well. In Mandarin, the alveolar fricative /s/ has major spectral prominence in the F4-F5 range, and the spectral noise band frequencies for /ʂ/ are centered between 2-3 kHz (Svantesson, 1986; Stevens *et al.*, 2004). Early works by Wu (1963) and Wu and Lin (1989) argued that the primary acoustic cue distinguishing Mandarin alveolars from retroflexes is the position of the lowest spectral prominence. In perception, Jeng (2009) reported that distinction between Mandarin alveolar and retroflex sibilants is correlated with the spectral center of gravity, a measure that captures the distribution of energy across the noise spectrum. Parallel to the English /s-ʃ/ spectral difference in Figure 1, Figure 2 displays the AI-grams of Mandarin /sa/ and /ʂa/ on the upper panel, as well as the AI-grams of Mandarin /su/ and /ʂu/ on the lower panel. The horizontal line on each AI-gram indicates the cut-off point of the noise distribution. The alveolars have higher cut-off frequencies than the retroflexes in both the /a/ and /u/ contexts. Moreover, the spectral distance (i.e., the distance between the two horizontal lines) between the alveolar and retroflex in the /a/ context is noticeably greater than that in the /u/ context (2.8 kHz between /sa/ and /ʂa/; 1.9 kHz between /su/ and /ʂu/), as the coarticulation from the following rounded vowel shrinks the spectral distance between /ʂ/ and /s/.

It should be noted that in addition to the spectral characteristics of frication, vocalic formant transitions have been reported to play a role in the identification of sibilants (e.g., Whalen, 1979 for English; Bladon *et al.*, 1987 for Shona). However, the contribution of formant transitions to fricative perception may be language-specific and depend on spectral similarity among target fricatives (Wagner *et al.*, 2006). In Mandarin, the formant transition cue was not found to significantly affect fricative perception (Chang, 2011b; Chiu, 2010). Chiu (2010) reported that Mandarin listeners were able to correctly identify /sa/ and /ʂa/ with incongruent formant transitions (e.g., /s/ followed by the vocalic portion crossed-spliced from /ʂa/). In Chang (2011b), it was found that Mandarin listeners' goodness ratings revealed no sensi-

[3] AI-gram (Li and Allen, 2011; Lobdell, 2009; Régnier and Allen, 2008) is a time-frequency representation. AI-gram integrates Fletcher and Galt's (1950) Articulation Index (AI) model of speech intelligibility and their critical-band auditory model, which can better evaluate the contribution of speech components to particularly consonant perception than fixed bandwidth spectrograms, especially in the presence of masking noise.

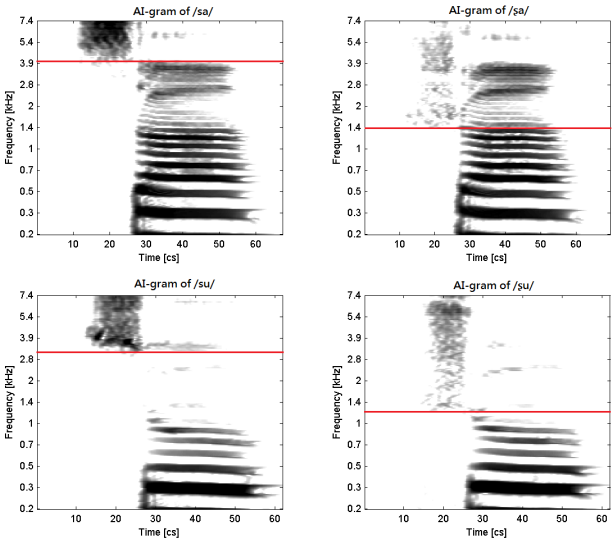


FIG. 2. AI-grams of /sa/ and /ʂa/ (upper panel); /su/ and /ʂu/ in Mandarin (lower panel).

tivity to mismatched transition in crossed-spliced /sa/ and /ʂa/ tokens. Therefore, in Mandarin, the cue contained in the frication noise is the most important cue in the alveolar-retroflex distinction.

D. Research questions

The goal of this study was to investigate whether speakers of a different dialectal background would perceive a phonological contrast differently in a given vowel context. Different acoustic realizations of the alveolar-retroflex contrast between Beijing and Taiwan Mandarin reported in previous research motivated a hypothesis that the category boundary will lie in different places along the spectral continuum. In addition, in Taiwan Mandarin, retroflex productions have been found to be acoustically less retroflexed than in Beijing Mandarin and the magnitude of the alveolar-retroflex contrast is smaller as well. Therefore, Taiwan Mandarin listeners were hypothesized to have different best category exemplars from their Beijing counterparts. Finally, given that both Beijing and Taiwan speakers produced a smaller place contrast in the /u/ context, the category boundary location and the internal structure were hypothesized to change across vowel contexts. Two experiments were designed to verify the aforementioned hypotheses. Experiment 1, an ABX task, was conducted to locate the category boundary between Mandarin alveolars and retroflexes. Experiment 2, a goodness rating task, investigated the internal structure of the /s/ and /ʂ/ phonemes to see if listeners judge certain exemplars of a phoneme category to be particularly good.

Four research questions will be addressed in verifying the aforementioned hypotheses: 1) Do Beijing and Taiwan Mandarin speakers share the same alveolar-retroflex boundary in perception? 2) Does the location of the category boundary vary across vowels? 3) Do Beijing and Taiwan Mandarin speakers share the same best exemplars for the alveolar and retroflex categories? 4) Does the location of a category's best exemplars vary across vowels?

III. METHODS

A. Participants

The participants for this study were 30 Beijing Mandarin (born and raised in Beijing or its vicinity) and 30 Taiwan Mandarin (born and raised in Taiwan) native speakers. They were between 18 and 35 years of age. All the Beijing participants speak only one dialect of Chinese, namely Beijing Mandarin, except for two who also speak the Nanjing and Shanxi dialects respectively. While half of the Taiwan participants speak some Taiwanese (a southern Chinese dialect), Mandarin was reported to be the major language used in their families. None of the participants self-reported any past or present speech or hearing disorders.

B. Considerations for stimuli construction

Only the /a, u/ vowel contexts were included for the stimulus construction in this study. As previously mentioned, the phonetic quality of /i/ is different when preceded by an alveolar vs. retroflex sibilant. In fact, in a continuum of /si-ʂi/, the two distinct allophones may facilitate the alveolar-retroflex distinction in perception (Lee, 2011). Therefore, vowel quality has to be manipulated if the /i/ context is to be included for the current study. Since it is not known what the impact of a phonologically non-existent vowel (i.e., a vowel between [ɪ] and [ɿ]) has on discrimination and goodness judgment of CV syllables, the /i/ context was not included in this study. The /ə/ context makes up very few alveolar-retroflex minimal pairs and therefore was not considered either.

Another factor that has been considered before constructing the stimuli was whether the stimuli should be produced by a single talker or multiple talkers. Literature on perceptual learning has pointed out that perception for fricatives is talker-specific (Eisner and McQueen, 2005; Kraljic and Samuel, 2005) and gender-specific (Mann and Repp, 1980). That is, stimuli produced by different talkers could result in quite different adjustments as listeners may maintain speaker-specific representations of each category. Experimentally, solving the problem of speaker normalization in speech perception introduces complications in data analysis. Therefore, the stimuli used in this study were produced by a single talker. As the production study by Chang and Shih (2012) reported that Beijing Mandarin speakers generally produced a larger alveolar-retroflex contrast than Taiwan Mandarin speakers, a Beijing Mandarin speaker was chosen to record the stimuli in order to maximize the acoustic contrast for subsequent /ʂ-s/ continuum construction.

C. Stimuli

As previously mentioned, the cue in distinguishing between Mandarin alveolar and retroflex sibilants lies in the frequency domain (/s/ has major spectral prominence above 4 kHz, and /ʂ/ centered around the F2 region). In light of the acoustic manipulations introduced in Li and colleagues (Li and Allen, 2011; Li *et al.*, 2012), once the cutoff noise frequencies (i.e., the frequencies of the lowest bound of the frication noise) of Mandarin /s/ and /ʂ/ are identified, raising the cut-off frequency of /ʂ/ to that of /s/ can naturally morph /ʂ/ into /s/. To create a continuum between /ʂ/ and /s/, one can edit out the frication noise with a high-pass filter, starting at the cut-off frequency of /ʂ/ in steps, until reaching the cut-off frequency of /s/. Beren, a software system developed by the Human Speech Recognition (HSR) Group at University of Illinois in Urbana-Champaign, allows such editing of naturally-produced speech signals using the short-time Fourier transform (Allen, 1977; Allen and Rabiner, 1977) and was used for constructing the stimuli for this study.

In obtaining naturally produced stimuli, a list of monosyllabic Mandarin words, which all carry tone 1, was spoken in isolation by a male native speaker of Beijing Mandarin. The words were presented in simplified Chinese characters to the talker on a computer screen in a randomized order, and each word was repeated 10 times. The recording took place in the sound-attenuated booth in the Phonetics Lab at the University of Illinois. At the end, one pair of /ʂa-sa/ and /ʂu-su/ that shared the same syllable duration (560 ms and 480 ms respectively) were extracted. These tokens were downsampled to 16000 Hz in compliance with Beren's requirement on sampling frequency, and were normalized for RMS amplitude at 65 dB. These tokens were read into Beren, where the AI-gram of each token was generated (as previously seen in Figure 2). To create an 8-step acoustic continuum from /ʂ/ to /s/ in each vowel context, the frication region between the alveolar and retroflex was divided into 7 equal intervals on the log scale, yielding 8 cutoff frequencies. See Table 1 for the cutoff frequencies of each step of the /ʂa-sa/ and /ʂu-su/ continua. The 8 steps of the continuum were constructed by modifying the frication in /ʂa/ and /ʂu/ at 8 cut-off frequencies. Note that the vowel portion was not modified in this process. Appendices 1 and 2 show how this step-by-step high-pass filtering of the noise was carried out for the /ʂa-sa/ continuum and the /ʂu-su/ continuum respectively.

All 8 steps of the continuum stimuli, including step 1 where the weak noise below the cutoff frequency was filtered out, were edited from the naturally-produced retroflex syllables. To verify the goodness of the edited sound files, they were played to two Beijing Mandarin and two Taiwan Mandarin native speakers. The sounds were judged to be natural sounding and the endpoints of the /ʂa-sa/ and /ʂu-su/ continua were unambiguously identified.

continuum steps	/ʂa-sa/ continuum	/ʂu-su/ continuum
Step1	1.39	1.24
Step 2	1.61	1.4
Step 3	1.89	1.61
Step 4	2.24	1.85
Step 5	2.67	2.12
Step 6	3.15	2.41
Step 7	3.68	2.72
Step 8	4.23	3.12

TABLE I. Cutoff frequencies (kHz) of each step of the /ʂa-sa/ and /ʂu-su/ continua (The cutoff frequency of each step was located using a logarithmic interpolation formula.)

In addition to the two fricative continua, two filler continua /i-y/ and /ti-di/ were also created, using the tokens produced by the same male Beijing Mandarin speaker. The /i-y/ continuum was created using the Akustyk package in Praat (Boersma and Weenink, 2010). The continuum varied by the F3 interpolated between /i/ and /y/. The /ti-di/ continuum varied by the duration of VOT and was created by splicing off the aspiration from /ti/ at equal time intervals into /di/.

The stimuli used in the ABX task consisted of two blocks. In each block, the B stimulus was two steps to the right of stimulus A on the continuum, making up six 2-step AB pairs from each 8-step continuum. Each AB pair was combined with stimulus X (i.e., stimulus A or B) and presented in four combinations (ABB, ABA, BAA, BAB), making a total number of 192 trials (4 acoustic continua * 6 two-step pairs * 4 presentation combinations * 2 blocks). All stimulus trials were presented in different random orders for each participant. The inter-stimulus interval within each ABX trial was set at 500 milliseconds.

The stimuli used in the goodness rating task came from the same acoustic continua described above. In this task, the participants heard a stimulus and saw the prompt question on screen asking "How is the pronunciation of this _____?", the blank being the character representation of either syllable from the minimal pair. For example, when step 1 (i.e., /ʂa/) from the /ʂa-sa/ continuum is played, the participants will be asked "How is the pronunciation of this SA?" in one trial, and "How is the pronunciation of this SHA?" in the other. This setup made no assumption that a bad pronunciation of /sa/ entails a good pronunciation of /ʂa/ (cf. Miller, 1994, where a stimulus from the /bi/-/pi/-excessively aspirated /pi/ continuum was only judged against /pi/, but not /bi/). That said, each step from the four continua was rated twice, making a total number of 64 trials.

D. Procedure

The participants were seated at a computer in a quiet room. They signed an informed consent form approved by the IRB at the University of Illinois. The data col-

lection session began with the ABX discrimination task, followed by a 5-minute break and the goodness rating task. The auditory stimuli were transmitted through earphones connected to the computer, and the participants made their responses by pressing the appropriate buttons on the keyboard (for the ABX task) and using a mouse (for the goodness rating task). The participants completed one questionnaire about their linguistic background after completion of the two tasks. The participants received cash compensation for their participation.

At the beginning of the ABX task, the participants were told that they would hear three Mandarin sounds (sounds 1, 2 and 3, which correspond to ABX) in a sequence and they had to judge whether sound 3 was the same as sound 1 or sound 2. A response page (see Appendix C) appeared on the monitor after each sound sequence was played. On the response page, the lines connecting sound 3 to sounds 1 and 2 respectively prompt the participants to press the key labeled 1 on the keyboard if they think sound 3 is identical to sound 1. Or the participants press the key labeled 2 on the keyboard if sound 3 is found to be identical to sound 2. All responses were logged automatically in E-Prime (v2.0; Psychological Software Tools, Pittsburgh, PA). The participants were encouraged to respond as quickly and accurately as possible. There was a 6-trial practice, and the participants had the opportunity to ask clarification questions before proceeding to the experiment.

In the goodness rating task, upon hearing a stimulus, a screen with a Visual Analogue Scale prompted the participants to rate the pronunciation (see Appendix D). With the mouse, the participants were asked to move the rhombus towards the left along the scale if they believed the sound they heard was a bad pronunciation of the prompted word, and towards the right if they believed that the sound was a good pronunciation of the prompted word. After they made their decision, two boxes appear below the scale asking them whether they would like to listen and rate again or proceed with the next sound. Participants were allowed to listen to each sound as many times as they deemed necessary. There were 16 trials in the practice. The practice stimuli included both ends of a continuum to implicitly familiarize the participants with the range of acoustic stimuli they would be hearing.

IV. RESULTS

A. ABX task

In categorical perception literature (e.g., Liberman *et al.*, 1957; Lisker and Abramson, 1967; Repp and Lin, 1989; Wood, 1976), listeners were found to best discriminate two sounds across the phoneme boundary. Following that, we assumed that a significant increase in accuracy in this task would be where the category boundary was located.

Figure 3 shows the plot of the mean /ʂa-sa/ discrimination accuracy with a 95% confidence interval for both listener groups. A one-way repeated measure ANOVA, with PAIR (i.e., six 2-step AB pairs) as the within-subject variable and ACCURACY as the dependent variable, was conducted on Beijing and Taiwan listeners' data respectively. For Beijing listeners, the analysis revealed a significant main effect for PAIR ($F(5, 145)=17.734$; $p<.001$), meaning that accuracy was not the same for all stimulus pairs. Post-hoc tests with a Bonferroni correction exploring the simple main effect of PAIR showed that pairs 1-3, 2-4 and 3-5 were not significantly different from each other, but pair 4-6 was significantly higher in accuracy than pairs 1-3, 2-4, and 3-5 ($p<.001$, .001, and .01 respectively). If listeners could not reliably differentiate step 3 from step 5 with a two-step distance in between, it is assumed that they could not perform better in differentiating between step 4 and step 5 with a one-step distance. On the other hand, listeners performed significantly better differentiating step 4 from step 6, suggesting that the phoneme boundary is between step 4 and step 6. Taken together, since the phoneme boundary is not between step 4 and step 5, it must be between 5 and 6, which corresponds to 2.91 kHz, halfway between the cutoff frequencies of steps 5 and 6 in the AI-grams (see Appendix 1 and Table I).

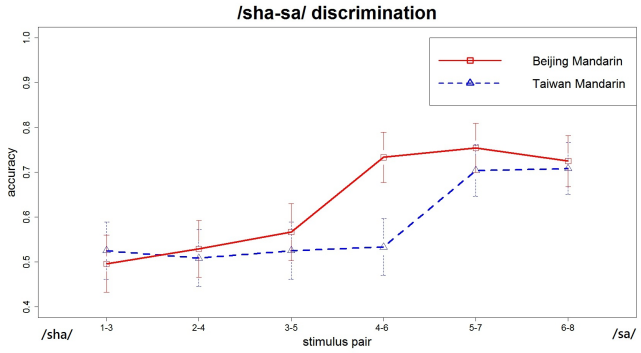


FIG. 3. /ʂa/-sa/ discrimination for Beijing and Taiwan Mandarin listeners (/sh/ stands for /ʂ/ in the plot).

For Taiwan listeners, the discrimination curve rises from a baseline value of approximately 0.5 accuracy to a peak of 0.75 at pair 5-7. The ANOVA analysis revealed a significant main effect for PAIR as well ($F(5, 145)=9.25$; $p<.01$). Pairwise comparisons showed that pairs 1-3, 2-4, 3-5 and 4-6 were not significantly different and the accuracy was at chance level, meaning that listeners could not discriminate among steps 1-6. On the other hand, the accuracy rate of pair 5-7 was significantly higher than pair 4-6 ($p=0.019$), which suggests that the perceptual discontinuity must have occurred between step 6 and step 7. This would correspond to 3.41 kHz, halfway between the cutoff frequencies of steps 6 and 7 in the AI-grams (see Appendix 1 and Table I).

The plot of the /ʂu-su/ discrimination performance for Beijing and Taiwan listeners is displayed in Figure 4. For Beijing listeners, a one-way repeated measure ANOVA analysis revealed a significant main effect for PAIR ($F(5, 145)=14.565$; $p<.001$). Post-hoc tests showed that pairs 1-3, 2-4, 3-5 and 4-6 were not significantly different from each other, but pair 5-7 was significantly higher in ac-

curacy than pairs 1-3, 2-4, 3-5 and 4-6 ($p<.001$, .001, .01, and .05 respectively). This suggests that the category boundary was located between steps 6 and 7, which corresponds to 2.56 kHz, halfway between the cutoff frequencies of steps 6 and 7 in the AI-grams (see Appendix 2 and Table 1).

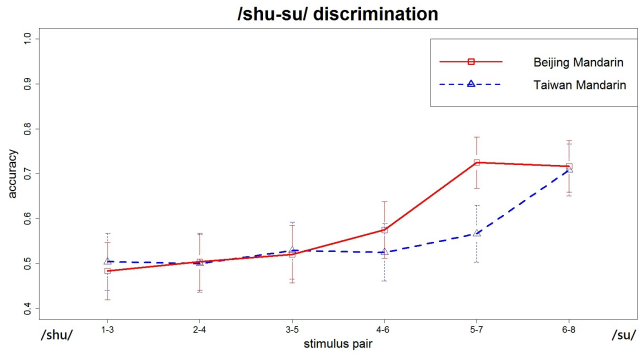


FIG. 4. /ʂu/-su/ discrimination for Beijing and Taiwan Mandarin listeners (/sh/ stands for /ʂ/ in the plot).

For Taiwan listeners, the ANOVA analysis also revealed a significant main effect for PAIR ($F(5, 145)=8.993$; $p<.01$). Post-hoc comparisons showed that only pair 6-8 was significantly higher in accuracy than pairs 1-3, 2-4, 3-5, 4-6 and 5-7 ($p<.001$, .001, .001, .001 and .05 respectively). That only pair 6-8 was discriminated above chance-level accuracy suggests that the category boundary was located between steps 7 and 8, since listeners could not discriminate among steps 1-7. This would correspond to 2.92 kHz, halfway between the cutoff noise frequencies of steps 7 and 8 in the AI-grams (see Appendix 2 and Table I).

B. Goodness rating task

As the /ʂ-s/ boundary location was decided based on the ABX data analysis, the two endpoints on the /ʂ-s/ continuum could be compared to their within-category variants (e.g., the endpoint /ʂ/ being compared to other stimuli on the continuum that are also considered /ʂ/). If an endpoint stimulus has significantly better rating scores than its sub-phonemic variants, then the endpoint stimulus is considered a better exemplar for its category. If the ratings for the sub-phonemic variants and endpoint stimulus are not significantly different, then the category is considered to be represented by an exemplar cloud without a better category member.

While participants could listen to the same stimulus and rate it as many times as they like in the goodness rating task, only the last response of each trial was used for data analysis. Since participants may have used the continuous scale differently—some may fully exploit the scale whereas others may be more conservative raters, all raw rating scores were transformed to z-scores. However, the plots using raw scores vs. standardized z-scores were similar. Therefore, raw scores were used for data plots and statistical analysis in the following analysis.

Figure 5 displays the plots of Beijing listeners' goodness judgments of stimuli from the /ʂa-sa/ continuum. The plot on the left shows listeners' ratings in response to the question "How is the pronunciation of this SHA?", and the plot on the right shows the ratings of the stimuli when being asked the opposite question "How is the pronunciation of this SA?" Each data point in the SA data set is inverted by 100 minus each point (the Visual Analogy Scale corresponds to a 0-100 scale) and correlated with its corresponding data in the SHA data. A strong correlation between the SHA dataset and the inverted SA dataset would suggest that listeners consider a bad instance of /ʂa/ a good instance of /sa/, and vice versa.

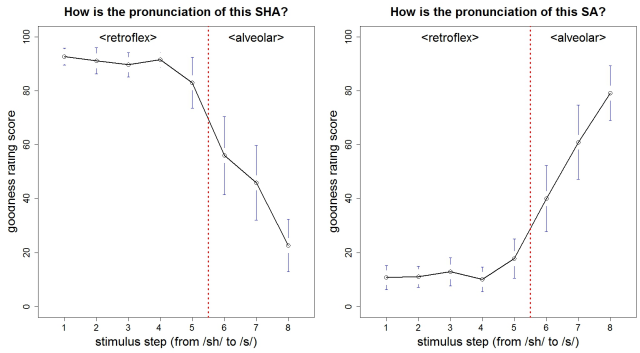


FIG. 5. Goodness rating of /ʂa/ and /sa/ along a /ʂa-sa/ continuum for Beijing listeners. The dotted line indicates the category boundary based on the ABX data.

The two sets of data were submitted to a one-way repeated measure ANOVA separately. For the SHA dataset (refer to the left panel in Figure 5), there was a significant main effect of STIMULUS ($F(7,203)=46.493$, $p<.001$), which means that goodness ratings were not the same across stimuli steps. Follow-up post-hoc tests revealed no statistical significance among steps 1-5 within the category of retroflex, suggesting that no single member of the retroflex category was rated better than the other. On the other hand, within the category of alveolar, step 8 was found to be significantly different from step 7 ($p=0.015$) and step 6 ($p<.001$). That is, step 8 was considered a better exemplar of the alveolar category. Now, the SA dataset (refer to the right panel in Figure 5) was submitted to the same statistical analysis and a significant main effect of STIMULUS ($F(7,203)=47.191$, $p<.001$) was also found. Follow-up post-hoc tests revealed no statistical significance among steps 1-5 within the category of retroflex. Within the alveolar category, step 8 was found to be significantly different from step 6 ($p<.001$) and marginally differently from step 7 ($p=0.055$). The correlation coefficient between the SHA dataset and inverted SA dataset was 0.797, suggesting the two datasets were strongly correlated.

For Taiwan listeners' goodness judgments on the /ʂa-sa/ continuum (Figure 6), a significant main effect of STIMULUS was found for both sets of data: $F(7,203)=10.11$, $p<.01$ for the SHA data set (refer to the left panel in Figure 6); $F(7,203)=11.825$, $p<.01$ for the

SA data set (refer to the right panel in Figure 6). Pairwise comparisons showed that no within-category variants were significantly different from their corresponding endpoint phonemes. The results suggest that for Taiwan Mandarin listeners, subphonemic variants of both /ʂa/ and /sa/ were perceptually equivalent to their respective endpoint stimulus. The correlation coefficient between the SHA dataset and inverted SA dataset was 0.401, suggesting the two datasets were moderately correlated.

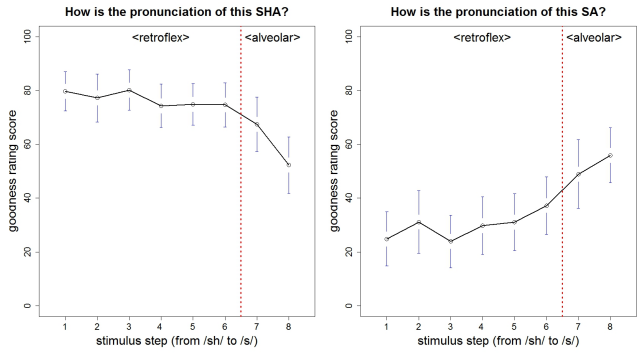


FIG. 6. Goodness rating of /ʂa/ and /sa/ along a /ʂa-sa/ continuum for Taiwan listeners. The dotted line indicates the category boundary based on the ABX data.

Figure 7 displays the plots of Beijing listeners' goodness judgments of the stimuli along the /ʂu-su/ continuum. A one-way repeated measure ANOVA revealed a significant main effect of STIMULUS for both sets of data: $F(7,203)=56.895$, $p<.001$ for the SHU data set (refer to the left panel in Figure 7); $F(7,203)=40.549$, $p<.001$ for the SU data set (refer to the right panel in Figure 7). For both datasets, pairwise comparisons showed no statistically significant difference among any steps within the category of retroflex, suggesting that all within-category members were considered equally good exemplars of /ʂu/. Within the category of alveolar, step 8 was found to be significantly different from step 7 in both sets of data ($p<.01$). The correlation coefficient between the SHU dataset and inverted SU dataset was 0.61, suggesting the two datasets were strongly correlated.

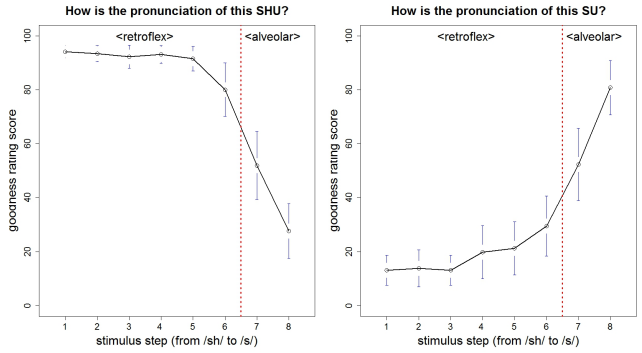


FIG. 7. Goodness rating of /ʂu/ and /su/ along a /ʂu-su/ continuum for Beijing listeners. The dotted line indicates the category boundary based on the ABX data.

Figure 8 shows Taiwan listeners' goodness judgments of the stimuli from the /ʂu-su/ continuum. A one-way repeated measure ANOVA statistical analysis revealed a significant main effect of STIMULUS for both sets of data: $F(7,203)=11.113$, $p<.01$ for the SHU dataset (refer to the left panel in Figure 8); $F(7,203)=16.856$, $p<.001$ for the SU dataset (refer to the right panel in Figure 8). For the SHU dataset, pairwise comparisons showed no statistically significant difference among any steps within the category of retroflex. For the SU data set, step 1 was only found significantly different from step 7 ($p<.01$) within the retroflex category. Since only step 8 was considered an alveolar based on the results of the ABX experiment, there was no other within-category member to compare to. Therefore, step 8 was the only alveolar exemplar in the stimuli being treated. The correlation coefficient between the SHU dataset and the inverted SU dataset was 0.467, suggesting the two datasets were moderately correlated.

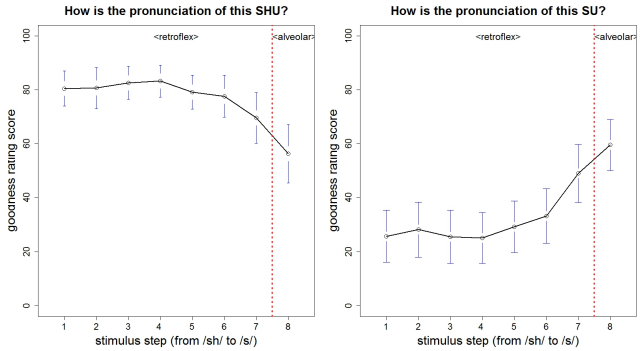


FIG. 8. Goodness rating of /ʂu/ and /su/ along a /ʂu-su/ continuum for Taiwan listeners. The dotted line indicates the category boundary based on the ABX data.

V. DISCUSSION

To locate the category boundary between Mandarin alveolar and retroflex sibilants, we conducted an ABX task where subjects discriminated stimuli along two acoustic continua that represent intermediate stages from the retroflex through the alveolar. The results showed that the continuous spectral properties were perceived categorically. We observed clear dialectal differences in the data: Beijing Mandarin and Taiwan Mandarin listeners had different perceptual boundaries along the /ʂ-s/ continua, with Taiwan listeners' perceptual boundary located one step closer to the alveolar end of the continuum than Beijing listeners. That is, given the same /ʂ-s/ continuum, Taiwan listeners allocated a wider range of noise frequency band to retroflexion than Beijing listeners. The location of the category boundary varied across vowels for both Beijing and Taiwan listeners: the /ʂ-s/ perceptual boundary shifted from 2.91/3.41 kHz in the /a/ context to 2.56/2.92 kHz in the /u/ context. This boundary shift indicates perceptual compensation for coarticulation from the following rounded vowel, in

line with Mann and Repp (1980) and Whalen (1979). To answer our first two research questions, a lower perceptual boundary frequency was found for Beijing listeners than for Taiwan listeners in both vowel contexts. This suggests that a lower cutoff noise frequency is required for the retroflex percepts for Beijing listeners, which may correspond to the variation in production where Beijing speakers generally produce a stronger degree of retroflexion than Taiwan speakers (Chang, 2011a; Chang and Shih, 2012). More steps along the continuum were perceived as retroflexes for both groups of listeners, suggesting that the /u/ context is prone to more perceptual confusion for the alveolar-retroflex distinction. This also well corresponds to a smaller separation between the alveolar and retroflex sibilants in the /u/ context in Beijing and Taiwan speakers' production (Chang and Shih, 2012).

We also conducted a goodness rating task to look into the internal structure of each phoneme and examine whether listeners would show gradient sensitivity to fine-grained acoustic differences in the spectral noise distribution. The results showed that Beijing and Taiwan Mandarin listeners perceived all variants of retroflexes as equally good (except for one instance noted in Taiwan listeners' SU data set). That is, both groups showed insignificant sensitivity or had great tolerance toward the sub-phonemic variation of the retroflex category. This pattern held true for both the /a/ and /u/ contexts. On the other hand, a gradient structure in the alveolar category was reflected in the goodness judgments of Beijing listeners for both vowel contexts. To answer the third and fourth research questions, both Beijing and Taiwan listeners judged the retroflex, in either vowel context, to be represented by an exemplar cloud without a best category member. Across vowel contexts, the best exemplar for the alveolar category was located in the alveolar end of the continuum for Beijing listeners. This was less clear for Taiwan listeners as there were few members (i.e., 2 members along the /ʂa-sa/ continuum and 1 member along the /ʂu-su/ continuum) in the alveolar category.

As we found more perceptual sensitivity to the acoustic variation within the alveolar category (refer to Figures 5-8), we noted that Lovitt and Allen (2006) have reported similar effects in the perception of English /s/ vs. /ʃ/. One question then arises of why there was discrepant sensitivity to within-category variation for the alveolar and the non-anterior fricative (i.e., retroflex, in the case of this study). An articulatory/acoustic account was provided here in an attempt to link between speech perception and speech production. Articulatorily, whether constriction is made behind the teeth or against the alveolar ridge, alveolars have a relatively small resonant cavity anterior to the constriction. Alveolars essentially involve only a tongue tip raising gesture. According to Keyser and Stevens (2006), "[b]efore a front vowel, the F2 starting frequency for an alveolar is only slightly higher than it is before a back vowel, indicating about the same fronted tongue-body position as that for the alveolar preceding a back vowel" (p. 48). The less variable articulatory gestures result in a more specific acoustic landmark for the alveolars—high-frequency frication noise. In contrast, retroflexion has more complex tongue configurations and articulatory properties (e.g., the subliminal space that adds volume and complexity to the resonant cavity), thereby contributing to more variability in its acoustics. In addition, optional enhancing gestures to the retroflexes like tongue blade raising (Stevens *et al.*, 2004) or lip rounding (Chang, 2010) can further contribute to the acoustic variability. Alternative support for the acoustic variability of the alveolar vs. retroflex productions may also come from Perkell and Nelson's (1985) production study on the vowels /i/ and /a/. They reported that a small change in tongue placement in the front cavity, as opposed to the back cavity, would create relatively large percentage changes in the area function at the constriction, which in turn would trigger more perceptually salient formant changes. Taken together, the acoustic properties associated with the retroflex consonant are more complex than those of the alveolar consonants (Stevens and Blumstein, 1975; p. 230), such that retroflex productions may exhibit greater variance than alveolar productions (as was found in Chang and Shih, 2012). We suggest that the acoustic characteristics may in turn impinge on perception such that listeners become more tolerant with a phoneme that has more acoustic variability, but more sensitive to a phoneme that has a more specific acoustic profile.

In studying perceptual best exemplars, Johnson *et al.* (1993) found that listeners' choices of perceptual best vowel exemplars to be more extreme (higher high vowels, lower low vowels, farther front front vowels, and farther back back vowels) than their own productions. They called this perceptual vowel space expansion a hyperspace effect. If the hyperspace effect extends to fricative perception, then Taiwan listeners are expected to give a better rating to retroflex stimuli produced by a Beijing speaker. However, this was not observed in their goodness rating data, as the rating scores on the endpoint /ʂ/ was not significantly different from those on the other /ʂ/ variants. Therefore, Taiwan listeners' retroflex perception data did not provide evidence for the hyperspace effect.

VI. FOLLOW-UP EXPERIMENT

In addition to discrepant within-category sensitivity described above, another rating discrepancy was observed for the /ʂ-s/ continuum from the data—listeners appeared more conservative with judgments on the alveolar members. Considering Beijing listeners' goodness rating scores on both /ʂa-sa/ and /ʂu-su/ continua for example (see Figures 5 and 7), it can be seen that step 1 /ʂu/ received an average of 90 points when rated against the SHU question and 10 points when rated against the SU question. However, when it comes to rating step 8 /su/, it received an average of 30 points in response to the SHU question and 80 points upon the SU question. Taiwan listeners' data also exhibited a similar pattern. This rating discrepancy may suggest that listeners perceived the retroflex members to be equally good, but were expecting even better alveolar exemplars than the /s/ endpoint stimulus. It is speculated that 1) the endpoint

alveolar stimuli were modified from the retroflex stimuli and may not be as good as naturally-produced alveolar stimuli. The modified /s/ stimuli contain partial spectral information (i.e., frication noise and CV transitions) from the retroflex stimuli, which might create a bias in favor of retroflex percepts. Alternatively, 2) a hyperspace effect (Johnson *et al.*, 1993) was induced in the ratings of alveolar members. The Beijing talker who recorded the stimuli produced an alveolar rather than dental /s/. It may be that listeners preferred more extreme /s/ productions, namely dental /s/, even though they may produce alveolar /s/ themselves. The first account essentially questions the validity of modification of retroflex stimuli to create the alveolar stimuli. The second account can only be tested by including the dental /s/ stimuli. To clarify the nature of this rating discrepancy, a follow-up experiment was conducted. Two questions were addressed: 1) Are the naturally-produced stimuli /sa/ and /su/ indistinguishable from the modified /sa/ and /su/ (i.e., step 8 of the 8-step /ʂ-s/ continua used in the ABX and goodness rating tasks)? 2) Are dental /s/ stimuli perceived as better exemplars of the alveolar category in Mandarin? To answer the first question, the goodness rating of the naturally-produced /s/ stimuli was compared to the modified ones, particularly, the step 8 stimuli. To answer the second question, two more steps that simulate the dental /s/ (by further raising the cutoff frequency of high-pass filters) were added to the /ʂa-sa/ and /ʂu-su/ continua.

Eight Beijing Mandarin speakers and eight Taiwan Mandarin speakers were recruited for the follow-up goodness rating experiment. The stimuli were 44 target stimuli (11 steps, including 10 steps of modified stimuli and 1 naturally-produced /s/, * 2 vowels * 2 prompt questions) and 16 filler stimuli (from the /ti-di/ and /i-y/ continua). The experiment was conducted following the same procedure introduced in Sec. III.

The data were submitted to a one-way repeated measure ANOVA. The statistical results can be summarized into two points: 1) The naturally produced /sa/ and /su/ were not scored significantly differently from the modified stimuli (i.e., step 8 from the continua used in the original experiment), and 2) the goodness ratings on steps 9 and 10 (i.e., the dental tokens) were not significantly different from Step 8 (i.e., the alveolar token).

Regarding the first point, we tabulated the goodness ratings of the naturally-produced /s/ vs. the modified /s/ for Beijing listeners and Taiwan listeners (see Appendix E). It can be seen that the modified stimuli generally had a slightly higher mean (although not at a significant level) and smaller variance than the corresponding naturally-produced stimuli. The results suggest that the low goodness ratings on the alveolars observed in Sec. IV were not the result of the existence of residue retroflex cues in the alveolar stimuli. Therefore, the experimental procedure using Beren to modify naturally-produced retroflexes by filtering out noise band is a valid method that creates natural-sounding alveolars.

The second point can be illustrated by plotting the goodness ratings on the 10-step continua (Appendices F and G). The dental tokens (steps 9 and 10) being rated equally good as step 8, an alveolar variant, indicates that Mandarin listeners did not prefer hearing dental /s/ to alveolar /s/. Therefore, the previously speculated hyperspace effect was not found. On the other hand, in accordance with the findings from our original experiment, both Beijing and Taiwan listeners judged the retroflex, in either vowel context, to be represented by an exemplar cloud without a best category member.

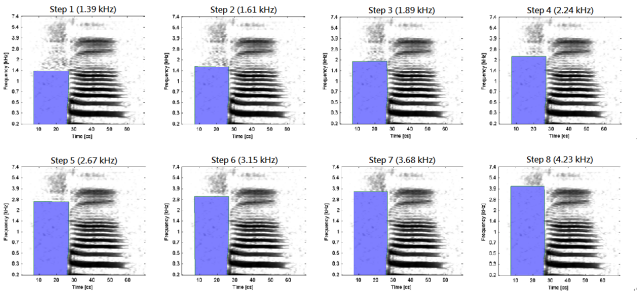
VII. CONCLUSION

This study shows that even when two dialects share the same repertoire of phonemic categories, there may still be considerable difference in the locations of category boundaries and phoneme exemplars. The two dialects, however, did converge in vowel context-moderated perception and lack of sensitivity to the sub-phonemic variation of the retroflex category. The cross-dialect perceptual patterns generally corresponds to variation in production, especially in terms of different thresholds for retroflex perception and confusion over the alveolar-retroflex contrast in the /u/ context. In conclusion, this study demonstrated the variability in cross-dialectal perception of the Mandarin alveolar-retroflex contrast. The categorical perception aspect of this study allows us to contribute to the body of literature on dialect effects in the perception of phonological contrasts. The selective gradient perception observed here sheds light on the variable relationship between acoustic variation and perceptual sensitivity as intermediated by articulation. Together, the results show that linguistic (i.e., vowel context) and sociolinguistic (i.e., dialect) factors collectively and variably affect the perception of phonological contrasts.

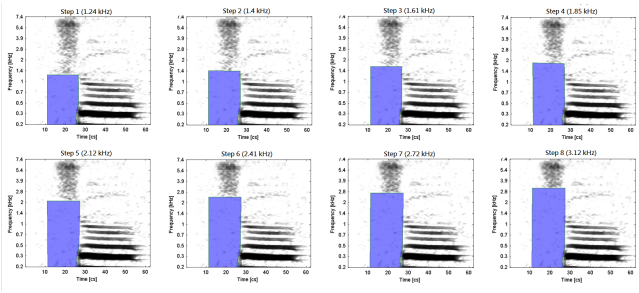
Acknowledgments

This research was supported in part by the NSF-funded project DHB: Fluency and the Dynamics of Second Language Acquisition (IIS-0623805).

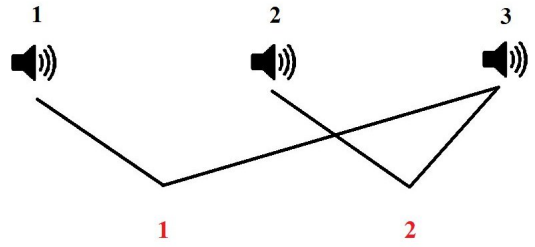
APPENDIX A: AI-GRAMS SHOWING THE CONSTRUCTION OF THE 8-STEP /sha-sa/ CONTINUUM (THE SHADED AREA INDICATES THE REGION OF NOISE THAT IS BEING FILTERED OUT.)



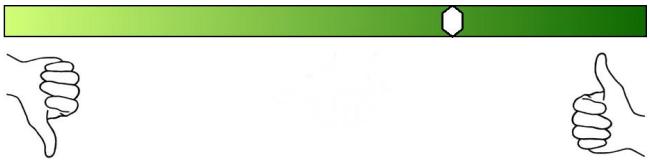
APPENDIX B: AI-GRAMS SHOWING THE CONSTRUCTION OF THE 8-STEP /shu-su/ CONTINUUM (THE SHADED AREA INDICATES THE REGION OF NOISE THAT IS BEING FILTERED OUT.)



APPENDIX C: A RESPONSE PAGE PROMPTING FOR DECISION IN THE ABX TASK



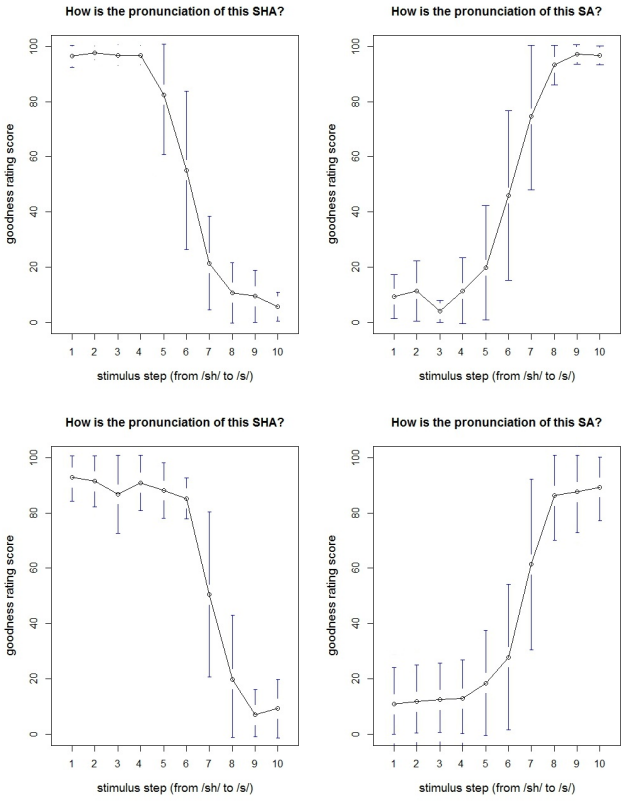
APPENDIX D: THE VISUAL ANALOG SCALE IN THE GOODNESS RATING TASK



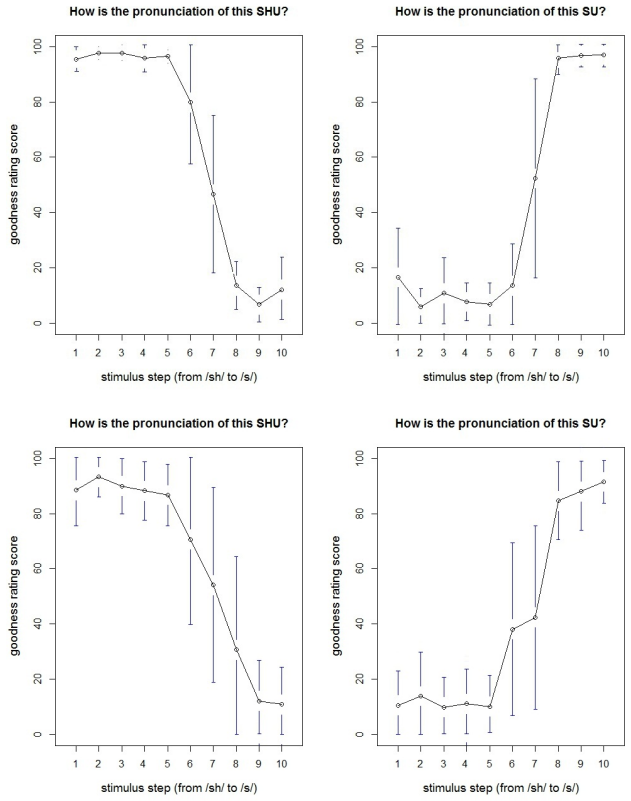
APPENDIX E: DESCRIPTIVE STATISTICS AND POST HOC ANALYSIS OF GOODNESS RATING OF NATURALLY-PRODUCED VS. MODIFIED TOKENS

dialect	stimulus	prompt question	mean	variance	pairwise comparison
Beijing	natural /sa/	SHA	10.75	170.214	p=.104
	modified /sa/		6.875	70.411	
	natural /sa/	SA	93.25	75.643	p=.42
	modified /sa/		96.125	23.268	
	natural /su/	SHU	13.5	150.268	p=.281
	modified /su/		7.85	89.554	
	natural /su/	SU	95.875	48.696	p=.436
	modified /su/		94.125	92.125	
Taiwan	natural /sa/	SHA	19.875	764.982	p=.299
	modified /sa/		12.25	363.929	
	natural /sa/	SA	86.25	387.071	p=.063
	modified /sa/		92.625	147.982	
	natural /su/	SHU	30.625	1643.411	p=.255
	modified /su/		23.75	980.214	
	natural /su/	SU	84.625	285.982	p=.302
	modified /su/		88.25	194.786	

APPENDIX F: GOODNESS RATING ALONG A 10-STEP /sha-sa/ CONTINUUM (STEP 8 IS THE ALVEOLAR /s/; STEPS 9 AND 10 ARE THE DENTAL /s/) FOR BEIJING LISTENERS (UPPER PANEL) AND TAIWAN LISTENERS (LOWER PANEL)



APPENDIX G: GOODNESS RATING ALONG A 10-STEP /shu-su/ CONTINUUM (STEP 8 IS THE ALVEOLAR /s/; STEPS 9 AND 10 ARE THE DENTAL /s/) FOR BEIJING LISTENERS (UPPER PANEL) AND TAIWAN LISTENERS (LOWER PANEL)



Allen, J. B. (**1977**). "Short-time spectral analysis, synthesis, with modifications, by discrete Fourier transform", IEEE Transactions on Acoustical Speech and Signal Processing **25**, 235–238.

Allen, J. B. and Rabiner, L. R. (**1977**). "A unified approach to short-time Fourier analysis and synthesis", Proceedings of the IEEE **65**, 1558–1564.

Allen, J. S. and Miller, J. L. (**2001**). "Contextual influences on the internal structure of phonetic categories: a distinction between lexical status and speaking rate", Perception & Psychophysics **63**, 798–810.

Bladon, A., Clark, C., and Mickey, K. (**1987**). "Production and perception of sibilant fricatives: Shona data", Journal of the International Phonetic Association **17**, 39–65.

Boersma, P. and Weenink, D. (**2010**). "Praat: Doing phonetics by computer", .

Brunelliére, A., Dufour, S., Nguyen, N., and Frauenfelder, U. (**2009**). "Behavioral and electrophysiological evidence for the impact of regional variation on phoneme perception", Cognition **111**, 390–396.

Chang, Y.-H. (**2010**). "Lip rounding in Taiwan Mandarin retroflex sibilants", Presented at the 87th Annual Meeting of the Linguistic Society of America, Baltimore, MD .

Chang, Y.-H. (**2011**a). "A corpus study of retroflex realizations in Beijing and Taiwan Mandarin", in *Proceedings of ICPhS XVII*.

Chang, Y.-H. (**2011**b). "The role of vowel transitions in the perception of Mandarin sibilants", in *Proceedings of the 162nd Meeting for the Acoustical Society of America*.

Chang, Y.-H. and Shih, C. (**2012**). "Using map tasks to investigate the effect of contrastive focus on the Mandarin alveolar-retroflex contrast", in *Proceedings of Speech Prosody 2012*.

Chao, Y.-R. (**1968**). *A grammar of spoken Chinese* (University of California Press, Berkeley).

Chien, C. G. (**1971**). "A contrastive study of the phonological systems of Mandarin and Taiwanese", Ph.D. Dissertation. Fu Jen University, Taipei, Taiwan.

Chiu, C. (**2010**). "Attentional weighting of Polish and Taiwanese Mandarin sibilant perception", in *Proceedings of the 2010 Canadian Linguistics Association Annual Conference*.

Chuang, Y.-Y. and Fon, J. (**2010**). "The effect of prosodic prominence on the realizations of voiceless dental and retroflex sibilants in Taiwan Mandarin spontaneous speech", in *Proceedings of Speech Prosody 2010*.

Chung, K. S. (**2006**). "Hypercorrection in Taiwan Mandarin", Journal of Asian Pacific Communication **16**, 197–214.

Clopper, C. G. and Pisoni, D. B. (**2004**). "Some acoustic cues for the perceptual categorization of American English regional dialects", Journal of Phonetics **32**, 110–140.

Cutler, A., Smits, R., and Cooper, N. (**2005**). "Vowel perception: Effects of non-native language vs. non-native dialect", Speech Communication **47**, 32–42.

Duanmu, S. (**2000**). *The phonology of Standard Chinese* (Oxford University Press, New York).

Eisner, F. and McQueen, J. M. (**2005**). "The specificity of perceptual learning in speech processing", Perception & Psychophysics **67**, 224–238.

Evans, B. G. and Iverson, P. (**2004**). "Vowel normalization for accent: An investigation of best exemplar locations in northern and southern British English sentences", Journal of Acoustic Society of America **115**, 352–361.

Evans, B. G. and Iverson, P. (**2007**). "Plasticity in vowel perception and production: a study of accent change in young adults", Journal of Acoustic Society of America **121**, 3814–3826.

Flanigan, B. O. and Norris, F. P. (**2000**). "Cross-dialectal comprehension as evidence for boundary mapping: Perceptions of the speech of southeastern Ohio", Language Variation and Change **12**, 175–201.

Fletcher, H. and Galt, R. (**1950**). "Perception of speech and its relation to telephony", Journal of Acoustical Society of America **22**, 89–151.

Harris, K. H. (**1958**). "Cues for discrimination of American English fricatives in spoken syllables", Language and Speech **1**, 1–7.

Heinz, J. M. and Stevens, K. N. (**1961**). "On the properties of voiceless fricative consonant", Journal of Acoustical Society of America **33**, 589–596.

Hughes, G. W. and Halle, M. (**1956**). "Spectral properties of fricative consonants", Journal of Acoustical Society of America **28**, 303–310.

Janson, T. and Schulman, R. (**1983**). "Non-distinctive features and their use", Journal of Linguistics **19**, 321–336.

Jeng, J.-Y. (**2006**). "The acoustic spectral characteristics of retroflexed fricatives and affricates in Taiwan Mandarin", Journal of Humanistic Studies **40**, 27–48.

Jeng, J.-Y. (**2009**). "The auditory discrimination of Mandarin retroflex contrasts and spectral moment analysis", Chinese Journal of Psychology **51**, 157–174.

Johnson, K., Flemming, E., and Wright, R. (**1993**). "The hyperspace effect: phonetic targets are hyperarticulated", Language **69**, 505–528.

Kawasaki, H., Ohala, J. J., and Jaeger, J. J. (**1986**). "Pho-

netic explanation for phonological universals: The case of distinctive vowel nasalization", in *Experimental phonology*, edited by J. J. Ohala and J. J. Jaeger, 81–103 (Academic Press, Orlando, FL).
Kendall, T. and Fridland, V. (**2012**). "Variation in perception and production of mid front vowels in the U.S. southern vowel shift", Journal of Phonetics **40**.
Keyser, S. J. and Stevens, K. N. (**2006**). "Enhancement and overlap in the speech chain", Language **82**, 33–62.
Kraljic, T., E., B. S., and Samuel, A. G. (**2008**). "Accommodating variation: dialects, idiolects, and speech processing", Cognition **107**, 54–81.
Kraljic, T. and Samuel, A. G. (**2005**). "Perceptual learning for speech: is there a return to normal", Cognitive Psychology **51**, 141–178.
Kratochvil, P. (**1968**). *The Chinese Language Today* (Hutchinson, London).
Kubler, C. C. (**1985**). "The influence of Southern Min on the Mandarin of Taiwan", Anthropological Linguistics **27**, 156–176.
Labov, W., Karan, M., and Miller, C. (**1991**). "Near-mergers and the suspension of phonemic contrast", Language Variation Change **3**, 33–74.
Ladefoged, P. and Maddieson, I. (**1996**). *The sound of the world's languages* (Blackwell Publishing, Oxford).
Lee, S. I. (**2011**). "An articulatory and acoustic investigation of Mandarin apical vowels", in *Proceedings of the 162nd Meeting of the Acoustical Society of America*.
Lee, W. S. and Zee, E. (**2003**). "Standard Chinese (Beijing)", Journal of the International Phonetic Association **33**, 109–112.
Li, F. and Allen, J. B. (**2011**). "Manipulation of consonants in natural speech", IEEE Trans. Audio, Speech and Language processing **19**, 496–504.
Li, F., Trevino, A., Menon, A., and Allen, J. B. (**2012**). "A psychoacoustic method for studying the necessary and sufficient perceptual cues of American English fricative consonants in noise", Journal of Acoustical Society of America **132**, 1–13.
Li, Y. S. (**2009**). "Incomplete neutralization of alveolar and retroflexed sibilants in Taiwan Mandarin", in *Proceedings of BLS 35*.
Liberman, A. M., Harris, K. S., Hoffman, H. S., and Griffith, B. C. (**1957**). "The discrimination of speech sounds within and across phoneme boundaries", Journal of Experimental Psychology **54**, 358–368.
Lin, Y. H. (**2007**). *The sounds of Chinese* (Cambridge University Press).
Lisker, L. and Abramson, A. S. (**1967**). "Some effects of context on voice onset time in English stops", Language and Speech **10**, 1–28.
Lobdell, B. E. (**2009**). "Models of human phone transcription in noise based on intelligibility predictors", Ph.D. Dissertation, University of Illinois at Urbana-Champaign.
Lovitt, A. and Allen, J. B. (**2006**). "50 years late: Repeating miller-nicely 1955", in *Proceedings of INTERSPEECH*.
Luo, C. and Wang, J. (**1981**). *Putong Yuyinxue Gangyao [Outline of General Phonetics]* (Shangwu Yinshuguan, Beijing).
Mann, V. A. and Repp, B. H. (**1980**). "Influence of vocalic context on the perception of the [sh]-[s] distinction", Perception & Psychophysics **28**, 213–228.
McMurray, B., Tanenhaus, M. K., and Aslin, R. N. (**2002**). "Gradient effects of within-category phonetic variation on lexical access", Cognition **86**, B33–B42.
Miller, J. L. (**1994**). "On the internal structure of phonetic categories: A progress report", Cognition **50**, 271–285.
Miller, J. L. and Volaitis, L. E. (**1989**). "Effect of speaking rate on the perceptual structure of a phonetic category", Perception & Psychophysics **46**, 1157–1167.
Niedzielski, N. (**1999**). "The effect of social information on the perception of sociolinguistic variables", Journal of Language and Social Psychology **18**, 62–85.
Perkell, J. S. and Nelson, W. L. (**1985**). "Variability in production of the vowels /i/ and /a/", Journal of Acoustic Society of America **77**, 1889–1995.
Pisoni, D. B. and Tash, J. (**1974**). "Reaction times to comparisons within and across phonetic categories", Perception & Psychophysics **15**, 285–290.
Preston, D. R., ed. (**1993**). *American Dialect Research* (Benjamins, Philadelphia).
Régnier, M. S. and Allen, J. B. (**2008**). "A method to identify noise-robust perceptual features: application for consonants", Journal of Acoustical Society of America **123**, 2801–2814.
Repp, B. H. and Lin, H. B. (**1989**). "Effects of preceding context on discrimination of voice onset times", Perception & Psychophysics **45**, 323–332.
Shadle, C. H. (**1985**). "The acoustics of fricative consonant", Ph.D. Dissertation. Massachusetts Institute of Technology.
Stevens, K. N. and Blumstein, S. E. (**1975**). "Quantal aspects of consonant production and perception: A study of retroflex stop consonants", Journal of Phonetics **3**, 215–233.
Stevens, K. N., Li, Z., Lee, C. Y., and Keyser, J. (**2004**). "A note on Mandarin fricatives and enhancement", in *From Traditional Phonology to Modern Speech Processing*, edited by G. Fant, H. Fujisaki, J. Cao, and Y. Xu, 393-403 (Foreign Language Teaching and Research Press, Beijing).
Sumner, M. and Samuel, A. G. (**2009**). "The effect of experience on the perception and representation of dialect variants", Journal of Memory and Language **60**, 487–501.
Svantesson, J. O. (**1986**). "Acoustic analysis of Chinese fricatives and affricates", Journal of Chinese Linguistics **14**, 53–70.
Volaitis, L. E. and Miller, J. L. (**1992**). "Phonetic prototypes: Influence of place of articulation and speaking rate on the internal structure of voicing categories", Journal of the Acoustical Society of America **92**, 723–735.
Wagner, A., Ernestus, M., and Cutler, A. (**2006**). "Formant transitions in fricative identification: The role of native fricative inventory", Journal of the Acoustical Society of America **120**, 2267–2277.
Wayland, S. C., Miller, J. L., and Volaitis, L. E. (**1994**). "The influence of sentential speaking rate on the internal structure of phonetic category", Journal of Acoustical Society of America **95**, 2694–2701.
Whalen, D. H. (**1979**). "Effect of vocalic formant transitions and vowel quality on the English [s]-[sh] boundary", Status Report on Speech Research, Haskins Laboratories **SR-59/60**, 35–48.
Wood, C. C. (**1976**). "Discriminability, response bias, and phoneme categories in discrimination of voice onset time", Journal of Acoustical Society of America **60**, 1381–1389.
Wu, Z. (**1963**). *The Collection of Illustrative Plates of Mandarin* (Commercial Press, Beijing).
Wu, Z. and Lin, M. (**1989**). *Overview of experimental phonetics* (Higher Eduction Press, Beijing).
Zhu, L. (**2012**). "Retroflex and non-retroflex merger in Shanghai accented Mandarin", M.A. Thesis, University of Washington.