USE OF SOCIAL INFORMATION IN THE PERCEPTION OF MANDARIN ALVEOLAR-RETROFLEX CONTRAST

Yung-hsiang Shawn Chang

Department of English, National Taipei University of Technology
shawnchang@ntut.edu.tw

ABSTRACT

Previous socioperceptual research has shown that speech perception can be affected by listeners' perceived dialect information of a speaker. While these studies have focused on vowels, the current study investigates whether similar dialect effects can be found in the perception of the Mandarin alveolar-retroflex contrast, a feature that is often used to distinguish Beijing and Taiwan Mandarin. The results show that Taiwan listeners' alveolar and retroflex identification judgments were not affected by the Taiwan/Beijing primes. On the other hand, the listeners have shown to be able to classify speakers by dialect above chance based on sentences and monosyllabic words. We discussed possible reasons for the null priming effects in relation to the indexical variation conveyed by consonants, as opposed to by vowels, and perceptual sensitivity to variants within the retroflex category. Data collection involving Beijing Mandarin listeners is underway.

Keywords: Mandarin alveolar-retroflex contrast, speech perception, dialect information

1. INTRODUCTION

Besides linguistic information, speech signals also contain social-indexical information such as the speaker's dialect background [11]. Both sources of information are encoded in memory and perceived by listeners. In fact, the two sources of information are intricately intertwined in perception: not only do the phonetic variants that are perceived affect what characteristics (e.g., gender, age, and dialect background) are attributed to a speaker; the characteristics attributed to a speaker can also influence how sounds are perceived. Several studies have shown that social information can interfere with linguistic processing. To take Niedzielski's [13] classic socioperceptual study for example, she examined the perception of the Northern Cities Chain Shift by asking listeners from Detroit, Michigan to match naturally-produced vowel stimuli to synthetically-produced vowel tokens (varying F1 and F2). Before the task, half of the listeners were told that the speaker was from Detroit and the other half were told that the speaker was from Canada. It was found that the listeners given the Canadian label chose raised-diphthong tokens, which more closely matched the speaker's actual productions, whereas the listeners in the Detroit-label group consistently selected the unshifted vowels as the best match. The results indicate that if listeners believe the speaker has certain characteristics, they may expect the speaker to sound a certain way and therefore be more likely to hear the speaker that way.

Other researchers further found that objects conveying the speaker's dialect information in the listener's environment can also affect how sounds are perceived. For example, Hay & Drager [7] covertly exposed their subjects to either stuffed toy kangaroos and koalas (associated with Australia) or kiwis (which are from New Zealand) prior to vowel perception task. As a result, the subjects' perception of the vowels shifted depending on which toys they had seen. Therefore, Hay & Drager concluded that the priming effect of dialect labeling is so robust that even implicit mention of another dialect would orient listeners' perception towards that dialect.

Most of these socioperceptual investigations have mainly focused on the perceptual integration of linguistic and social information conveyed by vowels. It may be that in languages like English, vowel differences across dialects outnumber consonant differences, and therefore are particularly important for distinguishing regional accents. In comparison, in Mandarin Chinese, vowels remain largely constant, whereas the repertoire of consonants varies across dialects. One consonantal feature that is most often used to characterize dialects of Mandarin is the distinction between the alveolar (/s, ts, tsʰ/) and retroflex (/ʂ, tʂ, tʂʰ/) sibilants. The retroflexes are replaced with their alveolar counterparts for Mandarin speakers who do not make the place distinction. Literature on Mandarin phonology (e.g., [5, 12]) generally prescribes a distinct place contrast made in Beijing Mandarin, whereas a lack or a lesser degree of distinction between the two categories is suggested for Taiwan Mandarin, another major dialect of Mandarin. In China and Taiwan, the use of retroflexion in speech is indexed for standard pronunciation and is associated with a higher education. As a result, speakers have been found to

exhibit stylistic retroflexing in speech. Jeng [9] reported that retroflexion occurs more often in response to a more formal register for Taiwan Mandarin speakers. Chung [4] pointed out that Taiwan Mandarin speakers are more subject to hypercorrect speech when they try to display respect or formality to the interlocutor. Given the social indexical properties of this place contrast, the question arose as to whether the priming effect of dialect labeling that was found for vowel perception is also present in the perception of consonants.

The presence of the priming effect of social information on perceiving a consonantal place contrast is premised on the fact that the perception of such a contrast is mediated by linguistic as well as sociolinguistic factors. In a recent study by Chang et al. [3], they found that Beijing and Taiwan listeners have different perceptual boundaries along the /ʂa-sa/ and /ʂu-su/ continua, with a lower cutoff frication frequency required for the retroflex percepts for Beijing listeners. That is, given the same retroflex-alveolar continua, fewer tokens were perceived as retroflexes by Beijing listeners. Both groups of listeners' category boundaries shift to lower frequencies in the rounded vowel context to normalize for vowel coarticulatory effects. These findings afford an investigation into whether manipulation of perceived dialect information affects the Mandarin alveolar-retroflex perception.

If the priming effect of regional labels is present following Niedzielski's [13] experimental paradigm, our Taiwan listeners' categorical boundary over the retroflex-alveolar continua is expected to shift toward the retroflex end of the acoustic continua when provided the Beijing label. One forced-choice identification task was specially designed to verify this prediction. We also included a forced-choice dialect categorization task to assure our listeners had the knowledge about dialect-specific variation in Mandarin, should null results for bias in perception were found.

2. METHODOLOGY

2.1. Participants

Sixty native Taiwan Mandarin speakers, born and raised in Taiwan, were recruited from the author's institution. They were between 18 and 25 years of age. While half of the Taiwan participants spoke 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.

2.2. Stimuli

The first task included four blocks of stimuli: an 8-step /ʂa-sa/ continuum, an 8-step /ʂu-su/ continuum, as well as two filler 8-step continua /i-y/ and /t^hi-ti/ (all carrying Tone 1) adopted from [3] (see [3] for details on the construction of these acoustic continua). It should be noted that the two retroflex-alveolar continua were constructed based on one Beijing Mandarin speaker's speech to maximize the acoustic distance between the alveolar and the retroflex, as Chang & Shih [2] reported that Beijing Mandarin speakers generally produced a larger alveolar-retroflex contrast than their Taiwanese counterparts. Each listener received the four blocks of stimuli in random order, and the stimuli were randomized within each block.

The stimuli for the forced-choice dialect classification task consisted of two successive blocks of stimuli: 4 alveolar- and 4 retroflex-initial monosyllabic words (respectively carrying each of the four Mandarin lexical tones), as well as the first two sentences of the Chinese version of *the North Wind and the Sun.* The stimuli were produced by eight Beijing Mandarin and eight Taiwan Mandarin speakers. The order of items was randomized within each block for each listener.

2.3. Procedure

The experiment was conducted in a sound-attenuated booth. The presentation and randomization of the stimuli were performed using E-Prime (v.2.0), and all responses were automatically logged. Prior to the experiment, the participants were randomly assigned to one of two groups: Taiwan-primed and Beijing-primed groups.

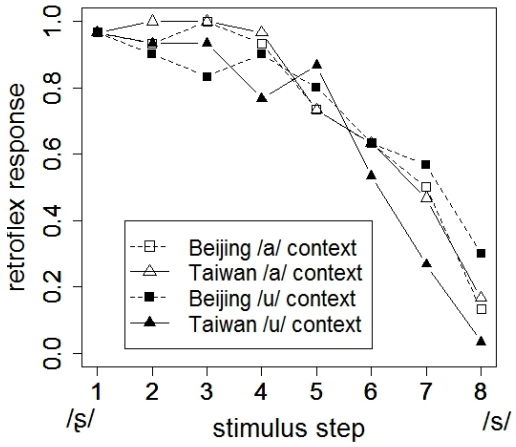
In Task 1, before each sound was played, a text page appeared on the monitor, instructing the participants that the speaker was from Taiwan (or Beijing, depending on the grouping assignment). Upon hearing the stimulus, listeners had to label the sound, by choosing one of the two characters representing either syllable from the minimal pair. Before the real task, there was a three-trial practice, with non-alveolar and non-retroflex stimuli produced by a different speaker.

In the first part of Task 2, listeners were instructed to judge where the speaker was from based on his or her pronunciation of the monosyllabic word that appeared on the screen. Each sound was played two seconds after the corresponding text page appeared. The second part of Task 2 followed the same procedure, except that the stimuli were sentences.

3. RESULTS

Figure 1 shows the plot of mean retroflex identification responses along the /ʂa-sa/ and /ʂu-su/ continua when primed with the Beijing vs. Taiwan label. Using the lme4 package [1] in the R software [15], we performed a mixed-effects logistic regression analysis where dialect label, vowel context and stimulus step were added as the fixed effects, and participants as a random intercept effect. The identification data was binary coded, and therefore logistic regression analysis was an appropriate statistical method. We started with an intercept-only model and then added the fixed factors one by one. The factors and factor interactions were retained if the model was significantly improved. The results revealed significant main effects of stimulus step (*Wald* $\chi^2(1)=360.37$, $p<.001$) and vowel context (*Wald* $\chi^2(1)=5.782$, $p<.05$), as well as significant interaction effects for stimulus step*dialect label (*Wald* $\chi^2(1)=6.251$, $p<.05$) and vowel context* dialect label (*Wald* $\chi^2(2)=7.628$, $p<.05$). Tukey-adjusted pairwise comparisons of estimated regression coefficients and z-statistics showed that Beijing-primed listeners did not perform statistically differently from Taiwan-primed listeners in identification along both /ʂa-sa/ and /ʂu-su/ continua. That is, the information of the speaker's dialect background was not found to affect Taiwan Mandarin listeners' categorization of the alveolar and retroflex stimuli in either vowel context.

Figure 1: Retroflex responses as a function of stimuli ranging from retroflexes to alveolars.



In classifying dialects based on sentences, participants' overall accuracy was above chance (M=92.5%; SD=12.4; t(59)=56.84; p<.001). Participants' overall accuracy in dialect classification based on alveolar- and retroflex-initial words was also significantly above the chance level of 50% (M=62.5%; SD=9.2; t(59)=-14.93; p<.001).

It should be noted that Fon & Chiang [6] found that tone production in Taiwan Mandarin is characterized by a narrower tonal range, lower tonal heights and flatter tonal contours than Beijing Mandarin. That is, differences in tonal realizations of monosyllabic stimuli may assist listeners in dialect classification. Kruskal-Wallis ANOVA, a non-parametric alternative to a one-way ANOVA, was used to determine whether listeners' classification performance differed by lexical tone of the stimuli, and a significant tone effect was found ($\chi^2(3)=19.75$, $p<.001$). The Nemenyi test for post-hoc comparisons showed that Tone 1 classification had significantly lower accuracy than the other three tones ($p<.05$, <0.5 and $<.01$ respectively), whereas the classification accuracy of Tones 2, 3 and 4 did not significantly differ from one another. Taken together, our participants were able to classify dialect based on sentences and alveolar and retroflex syllables above chance. Listeners performed worst in categorizing dialect based on words carrying Tone 1.

4. DISCUSSION AND CONCLUSION

The goal of this study was to investigate the role of dialect information in the perception of the Mandarin alveolar-retroflex contrast. We specifically asked whether the location of the perceptual boundary over the retroflex-alveolar continua shifts when given a different regional label. The results showed that Taiwan listeners' alveolar-retroflex identification judgments were not affected by the Taiwan/Beijing primes. We ruled out the possibility that our Taiwanese participants had no knowledge about the Beijing dialect, as they have shown to be able to classify dialect based on sentences and monosyllabic words above chance, although the accuracy was much lower for the latter. Recall in our analysis of dialect classification (Task 2) based on words, we found that lexical tones may be a factor that contributes to dialect categorization rather than, or in addition to, segmental information. More specifically, given that the fricative continua used in the identification task (Task 1) were alveolar- and retroflex-initial syllables carrying Tone 1, our listeners may not have been able to make their judgments with respect to segmental (i.e., consonantal) information alone. It follows naturally to ask if social indexical properties encoded in vowels are similarly encoded in fricatives. In the exemplar model of speech perception, every time a listener encounters a speech utterance, it is compared to previously stored exemplars, which are detailed with phonetic information and indexed with

social information (e.g., [10, 14]). The word-based exemplar model has been argued to involve phoneme-level access as well to account for the bias in vowel perception seen in [7] and [8]. In contrast to the null cross-dialect priming effect on the perception of Mandarin /s/ and /ʂ/, it should be noted that Strand [16] found that listeners' stereotypes about gender of the speaker biased their perception of English fricatives /s/ and /ʃ/. Whether dialect information is stored or represented qualitatively differently from gender in fricative exemplars should be further explored with empirical investigations.

Another explanation for the absence of effects of dialect labels on perception may have to do with listeners' lack of sensitivity to the within-category variation of the Mandarin retroflex, as reported in [3]. With a goodness rating task, Chang et al. found that both Beijing and Taiwan listeners were less sensitive to the retroflex variants than the alveolar counterparts. As they also reported that Beijing listeners' categorical boundaries were located a step closer to the retroflex end of the /ʂ-s/ continua than Taiwan listeners, the perceptual boundary of our Beijing-primed Taiwan listeners not shifting to the retroflex end may well be that they were not sensitive to the retroflex variants along the continua. Data collected from Beijing listeners, which is underway, may allow us to test this speculation, since Taiwan-primed Beijing listeners' perceptual boundary should move towards the alveolar end, where more within-category sensitivity was reported in [3], should the social effects be present.

Lastly, Niedzielski [13] attributed the priming effects of dialect information on speech perception to listeners' expectations and beliefs about a speaker. If Taiwan listeners consider the major differences between their own dialect and Beijing dialect lie in prosodic features (e.g., intonational and tonal realizations, speech rate), our identification task (Task 1), where the stimuli vary by fricative realizations, would not allow us to contribute to the discussion of social effects on perception. Therefore, additional research is needed with survey of listeners' knowledge about indexical variation in the other dialect.

5. ACKNOWLEDGEMENT

This study is supported by Ministry of Science and Technology, Taiwan, R.O.C. under Grant no. NSC 103-2410-H-027-003.

6. REFERENCES

[1] Bates, D., Maechler, M. 2010. *lme4 pack language in R* [software package], http://cran.r-project.org/web/packages/lme4/index.html

[2] Chang, Y.-H., Shih, C. 2012. Using map tasks to investigate the effect of contrastive focus on the Mandarin alveolar-retroflex contrast. *Proc. of Speech Prosody*, 302-305.

[3] Chang, Y.-H., Shih, C., Allen, J. B. 2013. Variability in cross-dialect perception of the Mandarin alveolar-retroflex contrast. *Proc. of International Conference of Phonetics of the Languages in China.*

[4] Chung, K. S. 2006. Hypercorrection in Taiwan Mandarin. *Journal of Asian Pacific Communication* 16, 197–214.

[5] Duanmu, S. 2000. *The phonology of standard Chinese*. New York: Oxford University Press.

[6] Fon, J., Chiang, W. Y. 1999. What does Chao have to say about tones? –a case study of Taiwan Mandarin. *Journal of Chinese Linguistics* 27, 15-37.

[7] Hay, J., Drager, K. 2010. Stuffed toys and speech perception. *Linguistics* 48, 865-892.

[8] Hay, J., Drager, K., Warren, P. 2010. Short-term exposure to one dialect affects processing of another. Stuffed toys and speech perception. *Language and Speech* 53, 447-471.

[9] Jeng, J.-Y. 2006. The acoustic spectral characteristics of retroflexed fricatives and affricates in Taiwan Mandarin. *Journal of Humanistic Studies* 40, 27-48.

[10] Johnson, K. 2006. Resonance in an exemplar-based lexicon: The emergence of social identity and phonology. *Journal of Phonetics* 34, 485–499.

[11] Ladefoged, P., Broadbent, D. E. 1957. Information conveyed by vowels. *J. Acoust. Soc. Am.* 29, 98-104.

[12] Lin, Y. H. 2007. *The sounds of Chinese*. Cambridge: Cambridge University Press.

[13] Niedzielski, N. 1999. The effect of social information on the perception of sociolinguistic variables. *Journal of Language and Social Psychology* 18, 62-85.

[14] Pierrehumbert, J. 2001. Exemplar dynamics: Word frequency, lenition and contrast. In: Bybee, J., Hopper, P. J. (eds), *Frequency Effects and Emergent Grammar*. Amsterdam: John Benjamins. 137–158.

[15] R Developement Core Team 2010. R: A language and environment for statistical computing. [computer program], http://www.R-project.org/

[16] Strand, E. A. 1999. Uncovering the role of gender stereotypes in speech perception. *Journal of Language and Social Psychology* 18, 86-100.