CROSS-DIALECTAL PERCEPTION OF VOICELESS DENTAL AND RETROFLEX SIBILANT VARIANTS IN TAIWAN MANDARIN

Yu-Ying Chuang & Janice Fon

Graduate Institute of Linguistics, National Taiwan University, Taiwan
d98142003@ntu.edu.tw; jfon@ntu.edu.tw

ABSTRACT

This study aims to investigate the perception of dental and retroflex sibilant variants in Taiwan Mandarin. Two listener groups were recruited and compared – the merged and the unmerged dialect groups. Results of the priming experiment showed that the two sibilant variants were perceived differently, primarily due to the varying frequency of occurrences. In addition, it was found that while the two dialect groups differed in their sibilant production, they were perceptually assimilated and both sibilant variations were processed similarly.

Keywords: dialect variation, speech perception, variant frequency, dental sibilant, retroflex sibilant

1. INTRODUCTION

Speech is full of variability, even within one single language. Within-language variability may come from individual speakers, or a group of speakers, i.e., dialects. As dialectal variation is frequently found in almost all languages in the world, a great number of production studies have been devoted to uncovering phonological variations among different dialects. Studies on how dialectal variations are perceived, however, come much later. Miller, et al. [6], for example, found that speakers of two French dialects utilize durational cues dissimilarly when perceiving vowels. Their results lead us to believe that speakers of different dialects also crucially differ in their perceptual systems. This has been evidenced by both behavioral [5] and neurological studies [1].

Despite having dialect-specific phonological mechanisms, speakers of different dialects are usually able to perceptually adjust themselves to dialectal variations. Such a perceptual adaptability is often attributed to the effect of experience, for cross-dialectal perception is not uncommon, given convenient transportation and media broadcasting in this modern era. In fact, it was found that different frequencies of contacting another dialect result in dissimilar perceptual consequences [7].

The effect of frequency is also a central topic when spoken word processing is investigated. In general, the more frequently a particular variant form occurs in one language, the more efficiently it enables listeners to access the lexicon. For instance, Connine [4] has studied the flapping rule of word-medial /t/ in American English. As it is a robust phonological rule, her results indicated that native listeners recognize target words with the flap pronunciation more easily than those with the unreduced [t^h] pronunciation. These findings are further taken as support for the exemplar-based account of stored lexical representations, which suggests that phonetic details are encoded for every encounter of a certain spoken word.

In light of the exemplar model, the perceptual relationship between standard forms and dialectal variant forms can be predicted. Specifically, for speakers of the standard dialect, the standard form should access the lexicon better than the variant form; speakers of the non-standard dialect, on the other hand, should process the variant form more easily, given that they are more accustomed to it. One good example to testify this prediction is to investigate the perception of retroflex variants in Taiwan Mandarin. There are three voiceless retroflex sibilants in Mandarin – /tʂ/, /tʂ^h/, and /ʂ/. As documented in the literature, the realizations of retroflex sibilants are dialectally determined [2]. In general, native northern dialect speakers usually preserve retroflex pronunciation (the unmerged group), whereas southern dialect speakers tend to replace retroflex sibilants with their dental sibilant counterparts – /ts/, /ts^h/, and /s/ (the merged group).

Aside from the production discrepancy between these two dialects, the linguistic environment also differs in northern and southern Taiwan. Taipei, the northern metropolitan city of Taiwan, is an immigrant city where a great number of citizens are from southern Taiwan. Owing to the population composition, both standard and variant forms of retroflex sibilants are frequently heard in

Taipei. On the contrary, deretroflexed pronunciation is dominant in southern Taiwan. If frequency does count, this external difference should presumably add variability to the processing of deretroflexed variants of speakers from the two dialects as well.

To complicate this issue further, not only deretroflexion, but also retroflexion, is observed in Taiwan Mandarin. That is, retroflex sibilants are substituted for their dental sibilant counterparts. Nonetheless, retroflexion is crucially dissimilar to deretorflexion in their frequency of occurrences. In particular, retroflexion is less often encountered, and it primarily occurs only in formal speech contexts [3]. As the variant frequencies differ, their lexical accessibility should not be the same, as predicted by the exemplar model.

To sum up, there are two specific research questions in the present study. First, how are sibilant variants in Taiwan Mandarin perceived by speakers of the two dialects, the merged and the unmerged groups? Second, given different variant frequencies, is there any discrepancy with respect to the processing of the deretroflexed and retroflexed variants? It is hoped that results of the current study could further our understanding towards cross-dialectal perception, and also shed lights on spoken word processing, especially when phonetic variations are involved.

2. METHOD

2.1. Participants

81 native Taiwan Mandarin speakers, aged form 18 to 25, participated in the experiment. None of them reported any history of hearing disorder.

2.2. Materials

Stimuli were 21 dental-initial bisyllabic words (e.g., *cong^{1}ming^{2}* 'smart') and 21 retroflex-initial bisyllabic words (e.g., *chao^{3}jia^{4}* 'quarrel'). These words were selected based on three criteria. First, no other dental or retroflex sibilants occur in the rest of the word. Second, for the first syllable of each word, the replacement of the initial sibilant with its counterpart does not lead to a tone-syllable gap (e.g., *zhi^{2}* could not be used since there is no *zi^{2}* in Mandarin). Third, the dental or retroflex counterpart of the stimulus is a nonword (e.g., both *chong^{1}ming^{2}* and *cao^{3}jia^{4}* are nonwords).

Given that the mutual substitution of dental and retroflex sibilants are existing phonological rules in Taiwan Mandarin, we attempt to include a control baseline in the experiment, and compare them with a non-existing rule, i.e., substituting aspirated stops ([p^{h}], [t^{h}], [k^{h}]) for unaspirated ones ([p], [t], [k]). Additional 21 bisyllabic words beginning with unaspirated stops (e.g., *bang^{1}mang^{2}* 'help') were selected. The three aforementioned criteria also applied during the selection of control stimuli.

The present experiment adopted a cross-modal form priming paradigm, in which participants were required to first listen to the auditorily presented prime and then perform a speeded lexical decision task on the target word visually shown on the computer screen. There were three priming conditions – standard, variant, and unrelated. The standard priming condition presented the canonical pronunciation of the target word; the variant condition presented the sibilant-substituted pronunciation; the unrelated condition was the presentation of another bisyllabic word that is neither phonologically nor semantically related to the target word. An example of the stimuli design is demonstrated in Table 1. All priming stimuli were recorded by a trained female phonetician, and were synthetically adjusted to 1000 ms in duration.

Table 1: Example stimuli (D: dental; R: retroflex; U: unaspirated).

	Prime			Target
	Standard	Variant	Unrelated	
D	*cong^{1}ming^{2}* 'smart'	*chong^{1}ming^{2}*	*fang^{1}xiang^{4}* 'direction'	*cong^{1}ming^{2}* 'smart'
R	*chao^{3}jia^{4}* 'quarrel'	*cao^{3}jia^{4}*	*fang^{2}wu^{1}* 'house'	*chao^{3}jia^{4}* 'quarrel'
U	*bang^{1}mang^{2}* 'help'	*pang^{1}mang^{2}*	*lang^{4}man^{4}* 'romantic'	*bang^{1}mang^{2}* 'help'

2.3. Procedure

For the form priming experiment, three stimulus lists were created. Every list contained only one priming condition of each target word, and the number of priming conditions was counterbalanced across the three lists. Participants were divided into three groups, and randomly assigned with one stimulus list. The presentation of the perception experiment was via E-Prime (version 1.1). Response time was recorded by a serial response box connected to the PC, measured from the onset of visual target presentation to subjects' response.

The perception experiment was followed by a production task. Given that our experiment was conducted in Taipei, where speakers are from various linguistic backgrounds, we could not make sure whether our subjects were native northern/

southern dialect speakers, or the descendents of the southern dialect speakers who moved to Taipei many years ago. With respect to the difficulty of identifying one's native dialect, our operational definition of dialect was based on speakers' own pronunciation, and we divided our participants into the sibilant merged and the unmerged groups, in resemblance of the southern and the northern dialects. The production task was a picture naming task, in which the production of six sibilant-initial words was elicited for each subject.

3. RESULTS

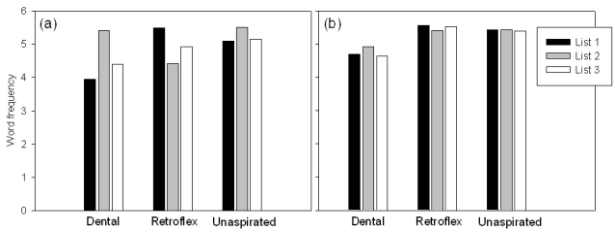
Two subjects were excluded from analyses, for they altered the responding fingers during the experiment. The separation of dialect groups was based on the auditory judgment of the first author, resulting in 31 subjects in the merged group and 48 in the unmerged group. Sibilant productions of both groups were acoustically measured. A two-way ANOVA with Dialect (merged/unmerged) and Sibilant (dental/retroflex) as factors was executed. There were significant effects of Sibilant [$F(1, 410) = 67.572$, $p < .001$] and Dialect × Sibilant [$F(1, 410) = 30.65$, $p < .001$]. Post hoc analyses showed that dental and retroflex sibilants were realized significantly differently for the unmerged group ($p < .001$), but not for the merged group.

To examine cross-dialectal perception of dental and retroflex sibilant variants, a three-way ANOVA with Dialect (merged/unmerged), Priming condition (standard/variant/unrelated), and Base (dental/retroflex/unaspirated) as between-subjects factors was performed. The main effects of Priming condition [$F(2, 4868) = 302.00$, $p < .001$] and Base [$F(2, 4868) = 3.16$, $p < .05$] were significant. The interaction of Priming condition × Base was significant as well [$F(4, 4868) = 14.65$, $p < .001$]. Post hoc analyses revealed that for both dental and retroflex sibilants, RTs of the standard and variant priming conditions were significantly shorter than those of the unrelated one, while the difference between them was not significant. As for unaspirated stops, it was shown that all pair-wise comparisons of the three priming conditions reached significance ($p < .001$).

The present perception results seemed to be self-explanatory, as we found that the variant forms of both dental and retroflex sibilants primed the intended target words as well as the standard pronunciations did, an effect that was absent when the phonological rule does not occur in the language (as in the case of unaspirated stop). Two expected effects, however, were missing. One was the dialect effect, for not any significant effect was reported for this factor. The other was the lack of processing difference between dental and retroflex sibilant variants. This was not expected, given their different variant frequencies.

It was therefore suspected that word frequency of our target words were not rigidly controlled. In this regard, 44 additional undergraduate students were recruited to rate the word frequency for each stimulus, on a scale from 1 (least frequently heard) to 7 (most frequently heard). Rating results are presented in Fig. 1a. As shown, frequencies varied across the three stimulus lists. In order to fix this problem, some target tokens were excluded from each list (5 for dental, 4 for retroflex, and 2 for unaspirated). Fig. 1b is the frequency data for the pruned sets of stimuli. For all three base types (dental/retroflex/unaspirated), word frequency did not significantly differ across the three lists.

Figure 1: The average word frequencies of three stimulus sets (a) before pruning (b) after pruning.

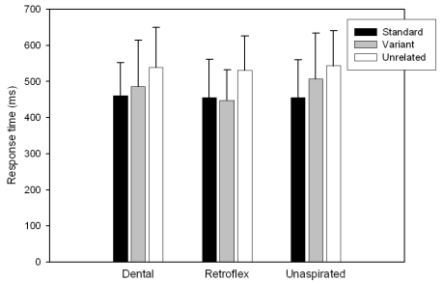


Also from Fig. 1b, we could see that target words with dental sibilants had overall lower word frequencies than the other two base types. To avoid unwanted confounding effects, the statistical test was run separately for each base type. A two-way ANOVA, with Dialect and Priming condition as between-subject variables, was performed. Results were comparable for all three base types: only the main effect of Priming condition was reported significant [dental: $F(2, 464) = 19.26$, $p < .001$; retroflex: $F(2, 697) = 53.69$, $p < .001$; unaspirated: $F(2, 1160) = 60.34$, $p < .001$]. Since Dialect was not significant at all, the pruned RT data of the two dialect groups are plotted together in Fig 2.

Results of unaspirated stops were similar to the original unpruned data. Post hoc comparisons reported that all three priming conditions differed from one another significantly ($p < .001$). What's dissimilar was the results of dental sibilants. Post hoc LSD tests showed that the variant condition not only differed significantly from the unrelated condition ($p < .001$), but also from the standard

condition ($p < .05$). Post hoc analyses of retroflex sibilants, on the other hand, did not report significant difference between the standard and variant priming conditions.

Figure 2: The response time (ms) of the three priming conditions for the three base types after data pruning.



4. DISCUSSION

The current study looked into cross-dialectal perception of sibilant variants in Taiwan Mandarin. The results of our priming experiment showed that Taiwan Mandarin listeners processed sibilant variant forms almost as efficiently as their standard forms, as compared with the control. This illustrated that both sibilant substitution rules were ubiquitous in Taiwan Mandarin, and therefore were pervasively adapted perceptually by listeners.

There still existed, however, subtle differences between the perception of the two sibilant variants. In general, deretroflexion was a more robust rule than retroflexion, as evidenced by the pruned data, showing that while both the standard and variant forms of retroflex sibilants primed the target word equally well, the priming effect of dental sibilants' variant forms was slightly weaker than that of standard forms. The result was compatible with the variant frequency. As predicted, since retroflexion less frequently occurs in Taiwan Mandarin, its lexical accessibility was thus weaker. Our finding also supported the exemplar-based account of lexical representations, in which frequency plays an important role in shaping our mental lexicon.

As for the effect of dialect, nonetheless, no statistical significance was achieved even after data were pruned. One possibility would be that the division of dialect groups was not authentic. As aforementioned, the two dialects are basically geographically bound. Since our experiment was conducted in Taipei, and since we did not control for how long our subjects had stayed in Taipei, it was possible that most of our subjects were already perceptually adjusted to the ambient speech environment of Taipei, where both standard and variant forms of retroflex sibilants abound. Different results could probably be obtained if this experiment were done in southern Taiwan. The other possibility was that the form priming paradigm was not sensitive enough for detecting the dialect effect. That is, if we believe that production systems do affect perceptual systems, dialectal differences should still be observed, even with our production-based division of dialects in the present experiment. It is therefore likely that other more sensitive paradigms would enable us to discover subtle dialectal differences.

In conclusion, how sibilant variants in Taiwan Mandarin are perceived by native listeners from the two dialect groups was empirically examined in the current study. A strong effect of variant frequency was found, indicating that more frequently occurred variants were more readily processed by listeners. Our results further served as a support for the exemplar-based account of lexical representations. The effect of dialect, nonetheless, was not detected in the present experiment. It was suspected that this might be due to the strong influence of the ambient speech environment, which assimilated subjects' perceptual mechanisms, despite their production differences. The other possible account would be the insensitivity of our experimental paradigm. For future studies, we shall keep pursuing the perception of sibilant variants in Taiwan Mandarin. We believe our findings will surely bring us further in understanding the complicated issue of cross-dialectal perception.

5. REFERENCES

[1] Brunellière, A., Dufour, S., Nguyen, N., Frauenfelder, U.H. 2009. Behavioral and electrophysiological evidence for the impact of regional variation on phoneme perception. *Cognition* 111, 390-396.

[2] Chuang, Y.Y. 2009. *An Acoustic Study on Voiceless Retroflex and Dental Sibilants in Taiwan Mandarin Spontaneous Speech.* Unpublished master's thesis. National Taiwan University, Taipei, Taiwan.

[3] Chung, K.S. 2006. Hypercorrection in Taiwan Mandarin. *Journal of Asian Pacific Communication* 16(2), 197-214.

[4] Connine, C.M. 2004. It's not what you hear but how often you hear it: On the neglected role of phonological variant frequency in auditory word recognition. *Psychonomic Bulletin & Review* 11(6), 1084-1089.

[5] Dufour, S., Nguyen, N., Frauenfelder, U.H. 2007. The perception of phonemic contrasts in a nonnative dialect. *J. Acoust. Soc. Am.* 121, EL131-EL136.

[6] Miller, J.L., Grosjean F. 1997. Dialect effects in vowel perception: The role of temporal information in French. *Language and Speech*. 40(3), 277-288.

[7] Sumner, M., Samuel, A.G. 2009. The effect of experience on the perception and representation of dialect variants. *Journal of Mermory and Language* 60, 487-501.