THE REPRESENTATION OF ONSET SIBILANT VARIANTS IN TAIWAN MANDARIN: AN ERP STUDY

Yun-Han Hsu, Janice Fon & Chia-Lin Lee

Graduate Institute of Linguistics, National Taiwan University, Taiwan
r09142006@ntu.edu.tw; jfon@ntu.edu.tw; chialinlee@ntu.edu.tw

ABSTRACT

Previous behavioural studies on Mandarin sibilant variants showed visual words are processed slower following their auditory forms of infrequent variants than frequent or canonical forms. This study further examined the processing timeline of variant frequency using the ERP technique. Forty-two native speakers viewed 120 sibilant-initial words following their auditorily presented canonical or variant pronunciations. Results showed larger N400 responses in the infrequent variants than the frequent ones during the auditory prime, suggesting greater effort in lexical access for the former. Responses to the subsequent visual words showed enhanced P2 for both variants relative to the canonical forms, but no such difference was found for N400 responses, indicating elevated attention following variants but similar semantic facilitation from the audio stimuli for all forms after a short delay. Our findings thus suggested the variant frequency effect emerges during lexical access prompted by variant pronunciation, and is quickly annihilated after visual presentation.

Keywords: speech perception, variant frequency, ERPs, retroflex sibilants, dental sibilants

1. INTRODUCTION

In Taiwan, aside from the official language, Taiwan Mandarin, other languages such as Min, Hakka, and Austronesian languages are used as well. Among them, Min is the second largest language and over 70% of the population in Taiwan are various degrees of Mandarin-Min bilinguals [1]. Due to the frequent contact between the two languages, the phonological system of Mandarin is inevitably affected by Min [2, 3]. In Mandarin, there are three retroflex sibilants (/tʂ/, /tʂ^h/, /ʂ/) and three dental sibilants (/ts/, /ts^h/, /s/). On the other hand, Min only has the dental series but not the retroflex ones. As a result of language contact, many Mandarin speakers show various degrees of merging between retroflex and dental sibilants [2, 3], resulting in deretroflexion of retroflexes [4]. This is a rather common variation, and could be found in both male and female speakers in casual speech [5]. However, males generally made smaller sibilant contrasts than females, but inconsistent cross-regional comparisons were observed, in which males from the southern regions made larger sibilant contrasts than ones from the northern regions [6]. Interestingly, there is also a reverse merging process reported, i.e., merging dentals with retroflexes, termed hypercorrection [7]. This likely occurs because speakers associate a negative connotation with dental realisations and thus substitute dentals with retroflexes even when the former is in fact the canonical form. Hypercorrection is mainly found in formal or read speech [7].

Previous behavioural studies showed inconsistent findings in the processing of pronunciation variants. Some studies claimed a canonical form advantage, which predicts that canonical forms are always processed more efficiently than variant forms. Studies on the medial /t/ deletion in English (e.g., *center* pronounced as *cenner*) showed that canonical forms (e.g., *center*) have advantages over variant ones (e.g., *cenner*) in processing even though /t/-deleted variants occur fairly frequently in that language [8]. However, other studies found that frequency plays an important role. High frequency pronunciations could facilitate spoken word processing even when they are not canonical. Sumner and Samuel [9] examined the r-less variant in the New York City (NYC) dialect and found that NYC residents process the frequent r-less variants faster than the canonical r-full forms. Results of Chuang and Fon [10] on Mandarin sibilants also sided more with Sumner and Samuel [9], and showed that visual words are processed slower following infrequent audio variants than frequent variants or canonical forms, while frequent variants are processed as quickly as the canonical forms.

Although frequency is found to be the major factor in variant processing, the intrinsic drawback of a behavioural study automatically prevents us from pinpointing when exactly such an effect occurs on the processing timeline, as reaction time is generally a reflection of a rather late stage of processing. It is thus unclear whether lexical activation is successful immediately after one hears the audio variant or some time is required before the correct lexical entry is accessed. This study thus examined when the variant frequency effect emerges during the early stages of processing using the event-related potential (ERP) techniques. By adopting a cross-model paradigm adapted from Chuang and Fon [10], participants first heard a word with one of the three sibilant variants, followed by a visual presentation of the word or an unrelated word to examine how variant frequency affects the early stage of processing. In order to make further comparison with Chuang and Fon [10] within a similar time course of the pronunciation variant processing, the interstimulus interval between the auditory and visual stimuli in this study was equivalent to the design in Chuang and Fon [10].

To examine how variant frequency affects the early stage of processing, we measured the three ERP components, the N400 time-locked to both auditory and visual stimuli, the visual P2, and the late positivity component (LPC) time-locked to the visual targets. The N400, a broad component that goes negatively and peaks around 400 ms post-stimuli, is usually associated with lexical access and semantic facilitation [11]. It also reflects the ease of retrieving phonological information at the lexical level. When more effort is required in the retrieval, larger N400 effects are elicited [12-14]. The visual P2 component, a positively going waveform that peaks at around 200 ms post-stimuli over anterior electrode sites, is known to be modulated by attention and is enhanced when the targets are relatively infrequent [15]. Finally, the LPC component reflects the explicit detection of semantic anomalies [16] and is thought to involve in conscious perception and recollection of an item [17]. For example, in a semantic congruency detection task, LPC was found to be larger for the unfamiliar dialectal pronunciations than for familiar ones [18]. We predicted that these ERP components (P2, N400, and LPC) would be modulated by the two different rules, the frequent deretroflexion rule [4] and the less frequent hypercorrection one [7].

2. METHOD

2.1. Participants

Forty-two native speakers of Taiwan Mandarin (21M, age=20-30) were recruited. None of them reported any history of hearing or neurological disorders.

2.2. Materials

Stimuli were 60 retroflex-initial (e.g., *she*^{4}*bei*^{4} 'equipment') and 60 dental-initial bisyllabic words (e.g., *zi*^{1}*xun*^{4} 'information'). Each word included three conditions: canonical, variant and unrelated pronunciations. Canonical forms were pronounced according to the dictionary pronunciation, while variant forms were pronounced by merging sibilants through deretroflexion or hypercorrection. Retroflexes were pronounced as dentals, while dentals were pronounced as retroflexes. Unrelated words were words that were not related to the sibilant-initial words in any way. An analogous set of 60 stop-initial pseudo variants were also included as a control baseline. This control baseline was designed to ensure that participants did not distinguish the variant pronunciations from the canonical ones based on merely similar phonological features they shared. In fact, only one phonological feature differed between canonical and variant pronunciations for the sibilant groups, [±retroflex]. Pseudo variants were thus created from these bisyllabic words by changing unaspirated stops to aspirated ones (e.g., *dian*^{4}*nao*^{3} 'computer' becomes '*tian*^{4}*nao*^{3}'), which was also one phonological feature apart. We expected that the maximum effect would be observed in the pseudo stop variants, compared to the sibilant groups. Table 1 is an example of the stimulus design. All priming stimuli were recorded by the third author, a female phonetician, and were later synthetically adjusted to one second using Praat [19].

	Auditory Prime			Visual Target
	Canonical	Variant	Unrelated	
R	*she*^{4}*bei*^{4} 'equipment'	*se*^{4}*bei*^{4}	*jie*^{2}*gou*^{4} 'structure'	*she*^{4}*bei*^{4} 'equipment'
D	*zi*^{1}*xun*^{4} 'information'	*zhi*^{1}*xun*^{4}	*guo*^{2}*nei*^{4} 'domestic'	*zi*^{1}*xun*^{4} 'information'
S	*dian*^{4}*nao*^{3} 'computer'	*tian*^{4}*nao*^{3}	*fa*^{1}*xian*^{4} 'discover'	*dian*^{4}*nao*^{3} 'computer'

Table 1: The stimulus design (R: retroflex; D: dental; S: stop).

2.3. Procedure

Three stimulus lists were created for the cross-modal experiment so that the same visual target occurred on a list only once. Word frequency was matched across lists. Participants were divided into three groups and were randomly assigned to one list. The experiment was presented via E-prime 3.0. Figure 1 illustrates the experimental procedure. To ensure that participants attend to the stimuli, they were asked to memorise both the auditory and visual stimuli presented in each block. A memory task was conducted right after the end of each block. Before the audio-visual experiment, a picture-naming task was conducted to record participants' production of the sibilants because the two sets of retroflex and dental sibilants are in free variation with each other and speakers often show various degrees of merging. Therefore, the picture-naming task was to check the speakers' degrees of merging and see whether there is an interaction between their production and perception.

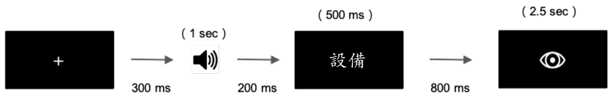

Figure 1: The flowchart of the experimental procedure.

2.4. EEG recordings

The electroencephalography (EEG) measurement was recorded using a 32-channel electrode cap via a *SYNAMPS2* amplifier with a 1000 Hz sampling rate. All electrodes were referenced to the left mastoids online and re-referenced to the average of the left and right mastoids offline. Four additional electrodes placed above/below the left eye and at the outer canthus of each eye in a bipolar montage measured the electrooculogram such as blinks and eye movements. Impedances were kept below 5 kΩ for all electrodes. The continuous EEG was segmented into epochs. Trials contaminated by artifacts were rejected before averaging. Artifact-free ERPs were averaged by stimulus type after subtraction of the 200 ms pre-stimulus baseline. Prior to measurement, ERPs were digitally filtered with a bandpass filter of 0.1-30 Hz.

3. RESULTS

For data time-locked to the onset of the second syllable of the auditory stimuli, a repeated measures ANOVAs with the within-subjects factor VARIANT (canonical/variant) was conducted for each sound on the mean amplitudes of the N400 responses measured between 300 and 600 ms in the central-posterior regions. The N400 responses on the retroflex showed no significant canonical-variant difference. For the dental and stop conditions, the N400 effects showed canonical-variant differences [dental: $F(1, 40) = 4.76, p < .05$; stop: $F(1, 40) = 13.72, p < .001$]. Results in Figure 2 showed that the N400 effect was larger in the variant condition than the canonical one in both dental and stop.

In addition, differences in mean voltage between the canonical and the variant condition were significantly different among the three sounds [$F(2, 80) = 6.00, p < .01$]. Post hoc analyses showed that the canonical-variant differences among three sounds were significantly different ($p < .001$). Figure 3 showed that the canonical-variant differences in stops was larger, followed by the ones in dentals and the differences were the least in retroflexes.

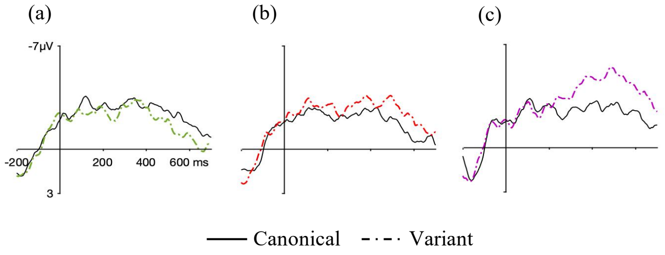

Figure 2: Grand average waveforms for the three conditions in the (a) retroflexes, (b) dentals and (c) stops, respectively, at electrode CZ when time-locked to the onset of the second syllable of the auditory stimuli.

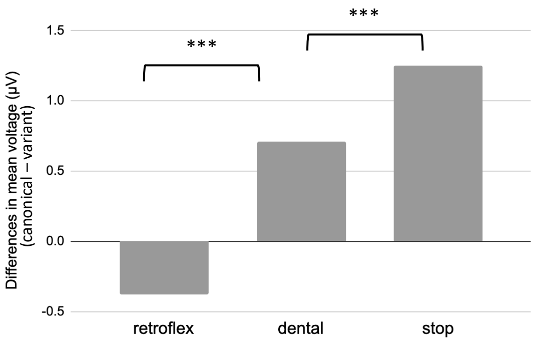

Figure 3: Canonical-variant differences in mean voltage between canonical and variant conditions when time-locked to the onset of the second syllable of the auditory stimuli (*** $p < .001$).

For data time-locked to the onset of the visual target, separate repeated measures ANOVAs with the within-subjects factor CONDITION (canonical/variant/unrelated) were conducted for each of the three components—the P2, N400, and LPC, for each sound. The P2 was computed within the 150-250 ms window in the frontal areas. The mean amplitudes of N400 were computed between 300 and 600 ms in the central-posterior regions. Lastly, The LPC was measured between 600 and 1000 ms in the central-posterior areas. Figure 4 respectively presents the waveforms time-locked to the onset of the visual stimuli in the (a) retroflexes, (b) dentals and (c) stops at electrodes FCZ, CZ, CPZ when time-locked to the onset of the visual stimuli.

For the P2 component, the main effect of CONDITION was significant among the three sounds [retroflex: $F(2, 82) = 6.20$, $p < .01$; dental: $F(2, 82) = 4.21$, $p < .05$; stop: $F(2, 82) = 4.25$, $p < .05$], with enhanced P2 effect in the variant condition across all three sounds (retroflex: $p < .001$; dental: $p < .01$; stop: $p < .001$), as shown in Figures 4(a), (b) and (c) at around 200 ms at electrode FZ.

Next, among all three sounds, the main factor CONDITION was significant in the N400 responses [retroflex: $F(2, 82) = 19.02$, $p < .001$; dental: $F(2, 82) = 15.21$, $p < .001$; stop: $F(2, 82) = 33.24$, $p < .001$]. Post hoc comparisons reported that the N400 effects were larger in the unrelated condition than the related condition across all sound types (*ps* < .001). However, the N400 canonical-variant difference was only significant in the stop group ($p < .001$), as shown in Figure 4(c).

The LPC responses showed that the conditional differences were only found in the stops [$F(2, 82) = 18.80$, $p < .001$]. Post hoc analyses showed that the LPC responses elicited larger responses in the variant condition than the other two conditions ($p < .001$), as shown in Figure 4(c).

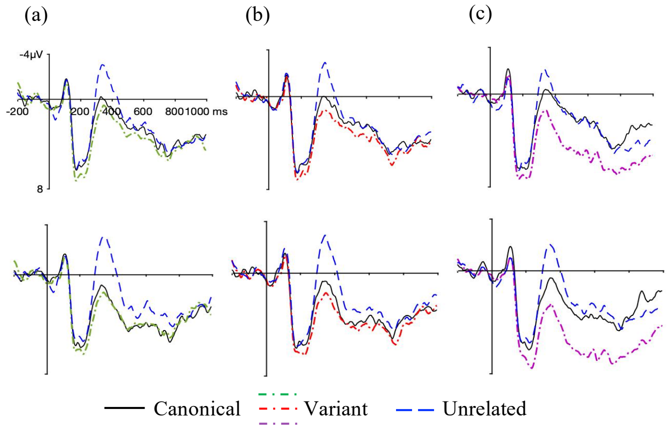

Figure 4: Grand average waveforms for the three conditions in the (a) retroflexes, (b) dentals and (c) stops, respectively, at electrodes FZ (upper row) and CZ (bottom row) when time-locked to the onset of the visual target.

Furthermore, based on the sibilant production, participants were divided into two groups, the unmerged (N=24, 10M) and the merged group (N=18, 11M). The unmerged group clearly distinguished retroflex from dental sounds while the merged group tended to replace the retroflex with dental sounds. No significant group differences were found in any of the factors.

4. DISCUSSION

This study examined the variant frequency effect of Taiwan Mandarin sibilant variants during the early processing stage. Results from the auditory stimuli showed that compared to the canonical conditions, the N400 responses were largest for the pseudo variants during auditory priming, followed by the infrequent variants. In contrast, no significant N400 difference was observed between the frequent variants and the canonical conditions. This suggests that, upon hearing the words, infrequent and pseudo variants required greater effort in lexical access than frequent ones, demonstrating a variant frequency effect in the early stage of processing.

When data was time-locked to the subsequent visual stimuli, responses showed enhanced P2 for both frequent and infrequent variants relative to the canonical forms, suggesting elevated attention. It indicated that regardless of variant frequency, variants were more "marked" compared to canonical pronunciations when a visual form was present, indicating a possible source for the canonical form advantage. Interestingly, no N400 difference was found between canonical and variant conditions, indicating similar semantic facilitation from the audio stimuli for both types of variants and their canonical forms after a short delay. In contrast to the data patterns from the existing frequent and infrequent variants, responses to visual words following the pseudo-variants showed enhanced P2 and LPC, reflecting attention and conscious recollection of the prime-target relationship. Our findings thus suggest that the variant frequency effect observed in the behavioural study by Chuang and Fon [10] emerges during the early stage of lexical access prompted by variant pronunciation. Moreover, this variant frequency effect is quickly annihilated after visual presentation, suggesting that the lexical access is successful in the later stage of processing. In addition, the null group effect suggested that there is a discrepancy between production and perception. Listeners can make a distinction between canonical and variant pronunciations even for the merged group. Put all the results together, our findings provided a possible reconciliation for the inconsistent patterns seen in previous behavioural literature.

5. REFERENCES

[1] S.-f. Huang, *Language, Society, and Group Awareness: Studies in Taiwan Sociolinguistics*. Taipei: Crane Publishing, 1994.

[2] C. C. Kubler, *The development of Mandarin in Taiwan: A case study of language contact*. Cornell University, 1981.

[3] C. C. Kubler, "The influence of Southern Min on the Mandarin of Taiwan,"

Anthropological Linguistics, vol. 27, no. 2, pp. 156-176, 1985.

[4] J. Fon, "The Phonetic Realizations of the Mandarin Phoneme Inventory: The Canonical and the Variants," in *Speech Perception, Production and Acquisition*: Springer, 2020, pp. 11-36.

[5] C. C. Lin, "A sociolinguistic study of the use of the retroflex sounds in Mandarin in College students in Taiwan," *Bulletin, College of Arts and Letters, National Central University,* pp. 1-15, 1983.

[6] Y. Chuang, "An acoustic study on voiceless retroflex and dental sibilants in Taiwan Mandarin spontaneous speech," *Unpublished master's thesis. National Taiwan University, Taipei, Taiwan,* 2009.

[7] K. S. Chung, "Hypercorrection in Taiwan Mandarin," *Journal of Asian Pacific Communication,* vol. 16, no. 2, pp. 197-214, 2006.

[8] M. A. Pitt, "The strength and time course of lexical activation of pronunciation variants," *Journal of Experimental Psychology: Human Perception and Performance,* vol. 35, no. 3, p. 896, 2009.

[9] M. Sumner and A. G. Samuel, "The effect of experience on the perception and representation of dialect variants," *Journal of memory and language,* vol. 60, no. 4, pp. 487-501, 2009.

[10] Y.-Y. Chuang and J. Fon, "Cross-dialectal Perception of Voiceless Dental and Retroflex Sibilant Variants in Taiwan Mandarin," in *ICPhS*, 2011, pp. 496-499.

[11] M. Kutas and K. D. Federmeier, "Thirty years and counting: Finding meaning in the N400 component of the event related brain potential (ERP)," *Annual review of psychology,* vol. 62, p. 621, 2011.

[12] N. Dumay, A. Benraiss, B. Barriol, C. Colin, M. Radeau, and M. Besson, "Behavioral and electrophysiological study of phonological priming between bisyllabic spoken words," *Journal of Cognitive Neuroscience,* vol. 13, no. 1, pp. 121-143, 2001.

[13] Y. Liu, H. Shu, and J. Wei, "Spoken word recognition in context: Evidence from Chinese ERP analyses," *Brain and language,* vol. 96, no. 1, pp. 37-48, 2006.

[14] A. S. Desroches, R. L. Newman, and M. F. Joanisse, "Investigating the time course of spoken word recognition: Electrophysiological evidence for the influences of phonological similarity," *Journal of Cognitive Neuroscience,* vol. 21, no. 10, pp. 1893-1906, 2009.

[15] S. J. Luck and S. A. Hillyard, "Electrophysiological correlates of feature analysis during visual search," *Psychophysiology,* vol. 31, no. 3, pp. 291-308, 1994.

[16] A. J. Sanford, H. Leuthold, J. Bohan, and A. J. Sanford, "Anomalies at the borderline of awareness: An ERP study," *Journal of Cognitive Neuroscience,* vol. 23, no. 3, pp. 514-523, 2011.

[17] M. Misra and P. J. Holcomb, "Event–related potential indices of masked repetition priming," *Psychophysiology,* vol. 40, no. 1, pp. 115-130, 2003.

[18] J. C. Bühler, F. Waßmann, D. Buser, F. Zumberi, and U. Maurer, "Neural processes associated with vocabulary and vowel-length differences in a dialect: An ERP study in pre-literate children," *Brain topography,* vol. 30, no. 5, pp. 610-628, 2017.

[19] P. Boersma and D. Weenink, "Praat: doing phonetics by computer [Computer program]. http://www.praat.org/ (accessed July 23, 2021)." 2021.