D1

How variations in syllable-final nasals affect word recognition in Taiwan Mandarin

Janice Fon, Yu-Ying Chuang (National Taiwan Univ.)
{jfon, d98142003}@ntu.edu.tw

1. Introduction

Human speech is full of variations. A simple word like *until*, the canonical (i.e., dictionary) form of which being /əntil/, can have a variety of pronunciations in daily conversation in addition to the prescribed form, including [əntʌ], [n̩tl̩], or even [tə] (Johnson, 2004). Therefore, how human listeners process spoken variants in their everyday lives is a topic that have intrigued researchers in the recent decade.

Interestingly, when facing the vast variability frequently encountered in speech production, human listeners do not seem to treat them in a uniform manner during spoken word processing. Instead, words in their canonical pronunciation seem to have an advantage and are processed more rapidly than words in non-canonical forms (e.g., Pitt, 2009; Ranbom & Connine, 2007; Sumner et al., 2014). Two accounts have been proposed for this processing bias. One explanation is to view human mental lexicon as being organized in a fashion similar to a dictionary, in which each lemma is phonologically coded in its canonical form only, and all its related variant forms are derived either via rules or associations (e.g., Gaskell & Marslen-Wilson, 1998; Gow, 2002). A canonical form thus enjoys processing efficiency not shared by its variant counterparts due to an exact mapping between the stored underlying form and the incoming speech signal. On the other hand, Sumner et al. (2014) argued that it is not the canonical status that makes a word form special, but rather, it is its social connotation that is granting it a distinctive place. They believe that the representation robustness of a word form is dependent on its social connotation. Word forms with positive connotation (i.e., more standard) are more strongly represented in the mental lexicon than words with negative connotation (i.e., less standard). In other words, the so-called "canonical form advantage" observed in many previous studies is in fact a "standard form advantage" instead, and the misconception comes from an alignment between canonical forms and standard forms that is frequently observed in languages.

This study thus intends to investigate the underlying cause for the spoken word processing bias using a particular set of syllable-final nasal mergers found in Taiwan Mandarin. This variation set is interesting because unlike most variations studied in previous research, the alignment between the canonical and the standard form is not always observed. There are in total three final nasal mergers, including /iŋ/→[in], /əŋ/→[ən], and /in/→[iŋ] (Fon et al., 2011). For /iŋ/→[in] (e.g., *gaoxing* 'happy' /kaʊɕiŋ/→[kaʊɕin]), the canonical [iŋ] is considered to be more standard than the variant [in] across dialects, showing perfect alignment. For /əŋ/→[ən] (e.g., *xueshen* 'student' /ɕɥeʂəŋ/→[ɕɥeʂən]), the canonical [əŋ] is considered to be more standard in some dialects, but the

variant [ən] is considered to be equally standard in others, showing only partial alignment. Finally, for /in/→[iŋ] (e.g., *aixin* 'a caring heart' /aɪɕin/→[aɪɕiŋ]), both the canonical [in] and the variant [iŋ] are considered to be equally standard across dialects, showing complete misalignment. In other words, one could consider the connotation of /iŋ/→[in] as mostly negative, that of /əŋ/→[ən] as relatively neutral, and that of /in/→[iŋ] as fairly positive. Therefore, if processing bias stems from differential social connotations of word forms [cf. Sumner et al. (2014)], then listeners should process the standard yet non-canonical [iŋ] from the /in/→[iŋ] merger, and, to a lesser extent, the standard-to-slightly-negative and non-canonical [ən] from the /əŋ/→[ən] merger with equal proficiency as their canonical counterparts of [in] and [əŋ], respectively. For the /iŋ/→[in] merger, listeners should still process the standard, canonical [iŋ] more efficiently than the nonstandard, non-canonical [in]. On the other hand, if processing bias is truly based on the canonical status of a word form [cf. Gaskell & Marslen-Wilson (1998); Gow (2002)], then listeners should only show a processing advantage for the canonical forms in all three mergers, regardless of its social connotations.

2. Method

2.1. Participants

A total of 198 native Taiwan Mandarin listeners, aged between 18 and 30, were recruited. Among them, 54 participated in Experiment 1, and 144 participated in Experiment 2, which had three sub-parts, one for each nasal merger, and each with 48 participants. None of the participants reported any language- or hearing-related disorders.

2.2. Materials

Experiment 1 included 45 bisyllabic stimuli, 15 for each nasal merger, and Experiment 2 included 90 bisyllabic stimuli, 30 for each nasal merger. These words were carefully selected so that both the canonical and the variant realization of the words are not homophonous with any other words in Mandarin. For example, both [kaʊɕiŋ] and [kaʊɕin] can only refer to *gaoxing* 'happy' /kaʊɕiŋ/, if at all. There were three prime conditions for each word, canonical, variant, and control. Canonical primes were realized with the dictionary form (e.g., [kaʊɕiŋ]), variant primes were realized with the appropriate variant form (e.g., [kaʊɕin]), and control primes were other bisyllabic words that were phonologically and semantically unrelated to the target words (e.g., *kafei* 'coffee'). For Experiment 2, there were also two target conditions for each word, canonical and variant. Table 1 shows an example of the prime and target conditions. For Experiment 1, all stimuli were recorded by the first author, who is a native speaker of Taiwan Mandarin. For Experiment 2, an additional male native speaker was also recruited to do the recording to avoid direct voice priming. Two lists were created via cross-splicing so that in each list, half of the trials were with male primes and female targets while the other half were with female primes and male targets. All stimuli were hand-adjusted to have the same duration using Cool Edit Pro (Version 2.0).

Table 1: An example of a stimulus *gaoxing* 'happy' /kaʊɕiŋ/ and its prime and target conditions.

	Prime (auditory)			Target		
	Canonical	Variant	Control			
Exp. 1	[kaʊɕiŋ]	[kaʊɕin]	*kafei* 'coffee'	高興		(visual)
Exp. 2	[kaʊɕiŋ]	[kaʊɕin]	*kafei* 'coffee'	Canonical:	[kaʊɕiŋ]	(auditory)
				Variant:	[kaʊɕin]	

2.3. Procedure

Both experiments were conducted in a sound-attenuated room. For Experiment 1, each trial began with an auditory prime, immediately followed by a visual target. For Experiment 2, an auditory prime was followed by an auditory target after a 500 ms pause. For both experiments, listeners were required to perform lexical decision on the visual/auditory targets as accurately and rapidly as possible. The experiment lasted about 20 and 40 minutes for Experiments 1 and 2, respectively, and listeners were compensated for their participation with a small monetary reward.

3. Results

Response time (RT) was measured from the onset of target presentation, and trials with incorrect responses (i.e., recognizing canonical or variant targets as nonwords) were excluded from RT analyses. For Experiment 1, a Prime (3) × Merger (3) ANOVA was performed (Figure 1). Results showed that only the main effects were significant [Prime: $F(2, 2372) = 116.33$, $p < .001$; Merger: $F(2, 2372) = 5.32$, $p < .01$]. Post hoc analyses indicated that control primes were always slower than the other two conditions ($p < .001$), but the difference between canonical and variant forms was not significant. RTs for /əŋ/→[ən] were overall longer than those for /iŋ/→[in] and /in/→[iŋ]. However, this was mainly contributed by the control primes, and exclusion of such removed the effect.

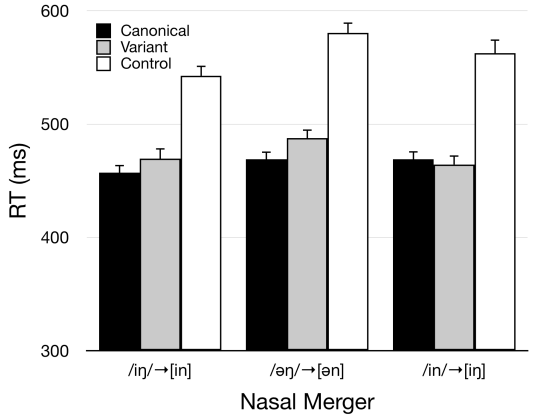

Figure 1: RTs of the three priming conditions in Experiment 1.

For Experiment 2, a Prime (3) × Target (2) × Merger (3) × Voice gender (2) ANOVA was performed. All main effects reached significance [Prime: $F(2, 8353) = 134.12$, $p < .001$; Target: $F(2, 8353) = 9.75$, $p < .05$; Merger: $F(2, 8353) = 19.56$, $p < .001$; Voice gender: $F(2, 8353) = 38.19$, $p < .001$]. Two interactions involving Target and Merger were also significant [Target × Merger: $F(2, 8353) = 18.67$, $p < .001$; Target × Merger × Voice gender: $F(2, 8353) = 3.09$, $p < .05$]. Post hoc analyses showed that the control primes were again with longer RTs than canonical and variant primes ($p < .001$), while the difference between the latter two was again not significant. For /iŋ/→[in], targets realized as variant forms (i.e., [in]) were recognized more slowly than those as canonical forms (i.e., [iŋ]) ($p<.001$) (Figure 2a). In contrast, for /in/→[iŋ], variant targets (i.e., [iŋ]) were more rapidly recognized than canonical targets (i.e., [in]) ($p < .005$) (Figure 2c). The patterns held true for both male and female voice. However, for /əŋ/→[ən], there was an additional interaction with voice gender. Variant targets (i.e., [ən]) were realized less efficiently than canonical ones (i.e., [əŋ]) only when they were uttered with the female voice ($p < .01$). The effect disappeared when the targets were in the male voice, and both canonical and variant targets were realized with comparable speed.

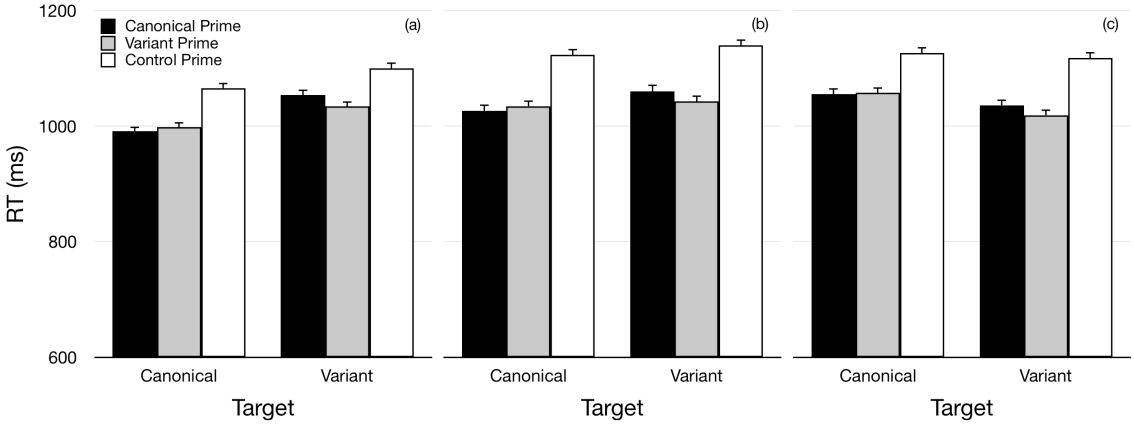

Figure 2: RTs for (a) /iŋ/→[in], (b) /əŋ/→[ən], and (c) /in/→[iŋ] in Experiment 2.

4. Discussion

This study examined the effect of a set of syllable-final nasal mergers on spoken word recognition in Taiwan Mandarin. Speech processing of canonical and variant forms of the three mergers, which differ in their social connotation, was compared using two form priming experiments. Results of Experiment 1 showed that the canonical form and its variant primed their linked written Chinese characters with comparable efficiency. This was rather surprising, as previous studies have shown otherwise [e.g., English: Ranbom & Connine (2007); French: Snoeren et al. (2008)]. Regardless of its origin, some kind of priming biases has always been observed. One suspects that this discrepancy might have been due to the peculiarity of the Chinese writing system, as its pronunciation encoding is not at all transparent. Thus, although written word recognition in alphabetic or syllabary writing inevitably involves phonological processing, this might not be necessarily so in Mandarin (Wu et al.,

2013). Compared to other languages, Mandarin users might be able to bypass detailed phonological processing when recognizing written word forms.

The uni-modal priming design in Experiment 2, on the other hand, revealed a significant effect of variation. Specifically, a robust effect of target realization was found on immediate spoken word recognition. Three different perceptual results were obtained. For /iŋ/→[in], the recognition of the canonical forms was more efficient than that of the variant forms. The processing advantage, however, greatly decreased for /əŋ/→[ən], since more efficient processing of the canonical forms only appeared in a voice-gender-dependent fashion. Critically, the result of /in/→[iŋ] was the most different, given that the processing of the variant forms outperformed that of the canonical forms. The current results were therefore in stark contrast to the hypothesis of the traditional account [cf. Gaskell & Marslen-Wilson (1998); Gow (2002)], because words encoded in canonical forms were not always more efficiently recognized than those in variant forms. They however strongly supported the social connotation hypothesis [cf. Sumner et al. (2014)], as the processing pattern was much in line with the connotation of the mergers. Based on the current findings, it can be seen that processing efficiency of a given form is not determined by whether it conforms to the canonical form. Instead, it is determined by whether this form is socially ideal or not. Previous findings of more robust processing for canonical forms thus likely stems from a frequent alignment between canonical forms and standard forms, rather than a true canonical form advantage in the mental lexicon.

However, similar to Experiment 1, priming difference between canonical and variant forms was still lacking in this experiment. This is surprising, as one would expect that same forms should prime better than different forms. However, as shown in Figure 2, one did observe a slight trend for a priming advantage between the same forms, and thus one suspects that the lack of significance might have come from the current experimental design. Given that there is a long interval of 500 ms between the offset of the primes and the onset of the targets, it is likely that any processing difference that might have existed between the two forms has already diminished or even disappeared by the time when auditory targets were presented. For future studies, inter-stimulus intervals could be manipulated in order to examine the time course of the priming effect in variant realizations.

5. Conclusion

By examining the three final nasal variations in Taiwan Mandarin, this study has thus demonstrated that it is social connotation that plays an important and decisive role in spoken word recognition, not canonical status. Socially ideal forms are more felicitous to lexical access than socially non-ideal ones, a finding much consistent with the socially weighted model proposed by Sumner et al. (2014). However, such an effect seems to be fairly short-lived and modality-sensitive. Unlike other languages [e.g., Ranbom & Connine (2007), Snoeren et al. (2008)], in which differential priming effects between canonical and variant forms have been consistently observed, even in cross-modal

experimental setups, no such evidence was found in the current study. Although one could probably attribute the null effect of differential priming from Experiment 1 to the logosyllabic nature of Chinese characters, the null effect of differential priming from the second experiment is not as easily explicable. Even though it is likely that the 500 ms inter-stimulus interval employed in this study is too long to detect a preferential priming effect, as mentioned above, Sumner & Samuel (2009) used exactly the same interval for their English participants and obtained significant results. It is thus possible that there might be cross-linguistic differences in spoken language processing with regards to the rate of dissipation in priming. It is also possible that different variations might have been the cause, as Sumner & Samuel (2009) examined syllable-final rhotics in English, and the current study looked at syllable-final nasals in Mandarin, which are acoustically less robust. More studies would be needed in order to understand the underlying cause for the preferential effects and the non-effects found in spoken variant processing.

References

Fon, J., Hung, J.-m., Huang, Y.-H., & Hsu, H.-j. (2011) "Dialectal variations on syllable-final nasal mergers in Taiwan Mandarin", Language and Linguistics 12(2), 273-311.

Gaskell, M. G. & Marslen-Wilson, W. D. (1998) "Mechanisms of phonological inference in speech perception", Journal of Experimental Psychology: Human Perception and Performance 24(2), 380-396.

Gow, D. W. (2002) "Does English coronal place assimilation create lexical ambiguity?", Journal of Experimental Psychology: Human Perception and Performance 28(1), 163-179.

Johnson, K. (2004) "Massive reduction in conversational American English." In K. Yoneyama and K. Maekawa (eds.) Spontaneous Speech: Data and Analysis. (pp. 29-54). Tokyo: The National International Institute for Japanese Language.

Pitt, M. A. (2009) "The strength and time course of lexical activation of pronunciation variants", Journal of Experimental Psychology: Human Perception and Performance 35(3), 896-910.

Ranbom, L. J., & Connine, C. M. (2007) "Lexical representation of phonological variation in spoken word recognition", Journal of Memory and Language 57(2), 273-298.

Snoeren, N. D., Seguí, J., & Hallé, P. A. (2008) "On the role of regular phonological variation in lexical access: Evidence from voice assimilation in French", Cognition 108(2), 512-521.

Sumner, M., Kim, S. K., King, E., & McGowan, K. B. (2014) "The socially weighted encoding of spoken words: a dual-route approach to speech perception", Frontiers in Psychology 4.

Sumner, M. & Samuel, A. G. (2009). "The effect of experience on the perception and representation of dialect variants", Journal of Memory and Language 60(4), 487-501.

Wu, J.-T., Yang, F.-L., & Lin, W.-C. (2013) "Beyond phonology matters in character recognition", Chinese Journal of Psychology 55(3), 289–318.