The Effect of Sibilant Merger on Phoneme Categorization in Taiwan Mandarin?

Vanessa Shih-Han Wu, Yu-An Lu

National Yang Ming Chiao Tung University

1   Introduction

The Ganong lexical effect has been extensively studied and demonstrates the relationship between the phonetic categorization of ambiguous signals and the influence of lexicality (Ganong, 1980). At its core, within a continuum where one end represents a word and the other end a non-word, participants demonstrate a strong bias towards identifying the word they heard. This phenomenon highlights the significant influence of top-down knowledge processing, particularly near the boundary of phonemes where ambiguity arises. However, it is important to note that lexical status is not the sole determining factor in the perceptual categorization process. Research has delved into various linguistic aspects to explore the impact of lexical effects. For instance, Connine et al. (1993) discovered the influence of word frequency on phonetic categorization. They observed that ambiguous tokens were more likely to be labelled as high-frequency words, reflecting an intrinsic alignment with lexicality. However, this tendency reversed when the influence of extrinsic factors was demonstrated; on the other hand, Politzer-Ahles et al. (2020) conducted research on Chinese participants and found a lack of frequency-based Ganong effect. However, a satisfactory explanation for this phenomenon is yet to be discovered. Furthermore, in a related study Yang et al. (2022) , the influence of suprasegmental factors, specifically tone frequency and markedness, on lexical effect was examined in Mandarin Chinese. Interestingly, they discovered an exception to the typical lexical effect, as T4 (X^{51}), despite being located at the gap end of the continuum, exhibited a stronger frequency effect compared to the word end T1 (X^{55}). Studies have indicated that the influence of lexicality can be surpassed by other linguistic factors, yet the exact understanding of each potential effect remains unclear.

Despite the widespread replication of studies on the Ganong effect in speech perception, most research has focused on factors that influence the mechanism of the lexical effect. However, there is limited evidence linking the dynamic aspect of sound change, particularly the merging of two phoneme categories, to listeners' identification. In the present study, the effect of sibilant merger on phoneme categorization in Taiwan Mandarin (TM) is investigated by conducting a two-alternative forced-choice identification experiment. Taiwanese variety of Mandarin has been well-documented that comparing to Standard Mandarin, the alveolar sibilants /s ts tsʰ/ and retroflex sibilants /ʂ tʂ tʂʰ/ undergo a process of merger primarily through the de-retroflexion of the retroflex category (Ing, 1984; Kubler, 1985). Lee-Kim & Chou (2022) examined the alveolar-retroflex sibilant merger in Taiwan Mandarin, namely the nature of this merger and discovered that the variation spans a full continuum, ranging from complete merger to clear contrast. The study focuses on the well-established process of merger between alveolar and retroflex sibilants in Taiwan Mandarin. It examines the interactions of various factors, including lexicality, frequency, and the influence of de-retroflexed variants of gap signals. Additionally, a control condition is included in the study, which consists of a continuum where both endpoints form words. This control condition allows for the observation of effects related to variants and frequency in the absence of a lexical effect.

2   Method

To investigate the categorization of sibilant mergers in Taiwan Mandarin, a two-alternative forced-choice identification experiment was carried out. A total of 126 stimuli were selected from 7-step continua

for each pair, resulting in a total of 18 pairs of continua. The selection of stimuli was based on the Centre of Gravity (CoG), for past researches such as (Jongman et al., 2000; Lee et al., 2014) considered CoG as one of the most reliable acoustic measures on sibilant fricatives. All items are listed in Appendix.

2.1 *Participants*   Ten Taiwan Mandarin speakers (5 female, 5 males, age 20-28, M = 22.9) were recruited from National Yang Ming Chiao Tung University. Experiment took place in Experimental Phonology Lab in the campus. None of the participants reported any speech or hearing disorders, and all participants were compensated monetarily for their time.

2.2 *Materials*   The stimuli consisted of 18 minimal pairs of disyllabic words with 6 targeted Mandarin sibilants in the word-initial position. Possible intrusive factors from the adjacent phonological environment sensitive to frication noise such as lip rounding is aware of and avoided as possible. These words were balanced for 3 pairs from different place of articulation ($ʂ – s$, $tʂ – ts$, $tʂ^h – ts^h$) and 3 continua with either a word on one end and a gap on the other, or words on both ends as the control group (gap – word, word – gap, word – word). Each continuum pair consists of two items.

    The stimuli were recorded by a female Taiwan Mandarin native speaker (aged 24). All token were scaled to 70 dB using Praat (Boersma & Weenink, 1992). In order to generate the sibilant merger continua, the initial sibilants in the words were excised and substituted with corresponded step-wise resynthesized segments. The sibilants were re-synthesized into 7 steps using Tandem-STRAIGHT with a gradual manipulation of their CoG values (Kawahara et al., 2008). Step 1 and Step 7 represented two endpoints across all continua, with Step 1 corresponding to the alveolar position and Step 7 corresponding to the retroflex position. **Table 1** shows the CoG values of the resynthesized stimuli for each merger pair measured from the spectrum using Praat's built-in function (Boersma & Weenink, 1992).

CoG (Hz)	$s – ʂ$	$ts – tʂ$	$ts^h – tʂ^h$
Step 1	9665.31	10353.02	9548.95
Step 7	5260.93	5049.67	6109.93
Range	4404.38	5303.35	3439.02

Table 1: CoG values of the resynthesized sibilants. The range of CoG values for Mandarin sibilants are calculated by subtracting the CoG value of Step 7 (retroflex position) from the CoG value of Step 1 (alveolar position). This calculation provides a measure of the overall span across the sibilant continuum.

2.3 *Procedure*   A total of 126 resynthesized stimuli [3 merger pairs (c – ch, s – sh, z – zh) × 3 continua (gap – word, word – gap, word – word) ×2 items × 7 steps] were presented using E-Prime (Schneider et al., 2012). The trials were randomized involving a six-trial practice to familiarize the participants with the positions of the label and task. The practice trials contained only the endpoints of stimuli from each all three targeted sibilant merger pairs. Participants were explicitly instructed to give the response based on the first sound of the first syllable they hear as fast as possible. The responses were recorded by pressing the corresponding key on the keyboard after they listened to each stimuli and judge whether the sibilants of the initial position they heard were alveolar (displayed on the left) or retroflex (displayed on the right). All stimuli were presented in Zhuyin, a non-alphabetical transliteration system commonly used in Taiwan Mandarin. The total duration of the experiment was around 15 min.

2.4 *Data analysis*   Generalized linear mixed-effect regression model was fit in R (*R Core Team (2020). — European Environment Agency*, n.d.) using the lme4 package(Bates et al., 2015) for statistical analysis. The dependent variable was Taiwan Mandarin speakers' sibilants response (retroflex position responses coded as 1, alveolar position responses coded as 0). The independent variables were step (scaled to centre), continua (gap – word, word – gap, and word – word; word – word as the baseline), pair ([$s – ʂ$] (coded as s – sh), [$ts – tʂ$] (coded as z – zh), and [$ts^h – tʂ^h$] (coded as c – ch)) and the interaction between continua*step to investigate the difference between gap – word and word – gap continua with word – word as the reference. The relative perceptual boundaries may vary depending on the specific merger pairs, but this variation is not relevant to the scope of the current study, the model did not include a three-way interaction among step-continua-pair. Random intercepts for participant and pair were included as well as by-participant random slopes for step, continua and by-pair random slopes for step and continua.

2.5 *Prediction*   The following predictions to be made: first, an effect of steps with a positive coefficient that the higher the step, the more retroflex (i.e. rightward) response should be given. Secondly, the lexical effect is anticipated to manifest across different continua, with the strongest impact observed at the gap end due to the ambiguous stimuli. It is anticipated that participants will provide more responses corresponding to the end of the continuum where words are formed. Furthermore, if there is an effect of de-retroflexion in Taiwan Mandarin sibilant mergers, a bias towards alveolar sibilants is expected. Specifically, with word – word continua as the reference, the lexical effect is expected to be stronger in the left response (i.e., the alveolar sibilant) of the gap – word continua, as alveolar sibilant gaps are more readily accepted as de-retroflexed variations of the retroflex sibilants. The participants are more likely to perceive the alveolar sibilant gaps in the gap - word continua as the de-retroflexion variants, indicating their inclination to provide more retroflex responses when encountered that position. Finally, based on the de-retroflexion of sibilant merger-in progress in Taiwan Mandarin, due to participants being explicitly instructed to focus on the onset of the first syllable rather than lexicality, a bias towards the left response is expected. (i.e., the alveolar side) in the word – word continua where no lexical effect is present.

3   Results

Figure 1 depicts the overall results of the experiment. The x-axis represents the scaled steps on a given sibilants continuum, while the y-axis represents the response, with values closer to 1 indicating more retroflex responses. The graph exhibits a clear positive coefficient between the steps and responses, suggesting that participants performed the task correctly. Consistent with the earlier prediction, a stronger lexical effect is expected in the alveolar sibilant response of the gap - word continua (GW, represented by the dashdotted line) compared to when it is located on the retroflex side (i.e., word - gap, WG, represented by the dashed line). **Figure 1** shows that between two continua gap – word and word – gap (i.e., the left side of GW vs. the right side of WG), in the gap - word continua, there is an earlier shift towards retroflex responses. This indicates that participants have a stronger inclination to provide retroflex responses when they encounter acoustically ambiguous stimuli where the retroflexed sibilant is associated with forming words, compared to when the alveolar sibilant is involved. Furthermore, the bias towards the alveolar end in the word-word continuum, where no lexical effect is present, can be seen as evidence of an ongoing merger of the sibilants towards the alveolar end. Therefore, when considering the randomized nature of all trials, participants tend to perceive a higher proportion of alveolar sibilants compared to retroflex sibilants when both ends of the continuum form words.
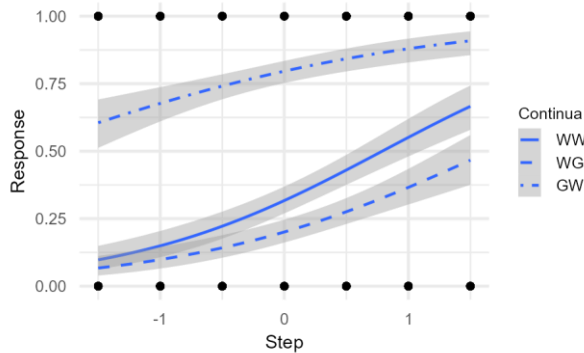

Figure 1: The aggregate results of estimated proportion of retroflex sibilants responses as a function of scaled step (centred as 0) and 3 continua.

As summarized in **Table 2**, the statistical results of the GLMM model, with the word–word continua as the reference level, revealed a significant step effect with a positive estimate ($\beta = 1.20$, $p < .001$). This finding suggests that retroflex responses increased as participants heard stimuli with lower CoG values and accurately identified them. Significant continua effects were found in both gap – word (GW) continua ($\beta = 2.99$, $p < .001$) and word – gap (WG) continua ($\beta = -1.68$, $p = .038$) between word – word continua (reference level) confirmed the present of Ganong lexicality effect in which the participants gave different responses based on the presence of ends in both continua. Specifically, more responses were given towards to the end that makes words (i.e. positive estimate in GW continua) and avoided gap ends across all continua (i.e., negative estimate

in WG continua). The findings further confirmed the visual observation that there is a stronger inclination towards retroflex responses when the alveolar end is presented as an acoustically ambiguous stimulus (i.e., GW), compared to its retroflex gap counterpart (i.e., WG) based on a stronger effect size on the estimated coefficients. No significant interaction between step*continua was observed, suggesting that the relationship between step and response doesn't significantly differ between WW and both GW and WG continua.

WW as the reference					
	Est.	*SE*	*z*	*P*	
(Intercept)	-0.863	0.52	-1.65	.099	.
step	1.207	0.29	4.16	<.001	***
WG	-1.682	0.81	-2.08	.038	*
GW	2.997	0.75	3.98	<.001	***
step: WG	-0.146	0.26	-0.56	.579	
step: GW	-0.336	0.23	-1.44	.150	

Table 2: The summary of fixed effects in the model glmer [Response ~ Step * Continua + (1 + Step + Continua | Subject) + (1 + Step + Continua | Pair), family = binomial]. (WG as word-gap continua, GW as gap-word continua, and WW as word-word continua)
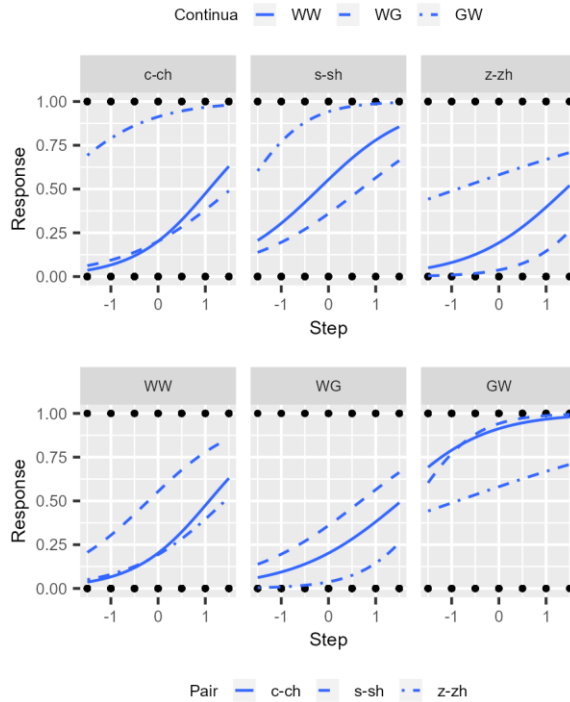

Figure 2: Result from the experiments by continua (top row) and by pair (bottom row). The x-axis represents the scaled steps on a given sibilants continuum; the y-axis represents the response, with values closer to 1 indicating more retroflex responses.

Figure 2 further depicts the results of the experiment by continua and by pair. From the graphs depicting the different continua (top row), an overall pattern emerges that is shared across pairs. The word-word continua (WW) falls in between the gap-word continua (GW) and the word-gap continua (WG), where the lexicality effect is involved. Additionally, it's observed that s-sh [s–ʂ] pair, comparing to c-ch [tsʰ–tʂʰ] and z-zh [ts–tʂ], demonstrates more balanced pattern in the word-word continua, where there is an increase in retroflex responses corresponding to the steps. On the other hand, c-ch [tsʰ – tʂʰ] and z-zh [ts – tʂ] pairs exhibit an asymmetrical pattern with a bias towards the alveolar endpoints, indicating the de-retroflexion merger status in Taiwan Mandarin. It is worth noting that participants were explicitly instructed to focus on the onset of the first syllable rather than lexicality. As previously predicted, a bias towards the left response is expected in randomized trials. Finally, it is worth noting the observed resistance of retroflex responses in the z-zh [ts - tʂ] pair compared to the other pairs in the gap-word (GW) continua. Conversely, the c-ch [tsʰ-tʂʰ] pair demonstrates an opposite pattern. The aforementioned topics of concern have been addressed and included in the discussions that follow.

4　Discussion

In a two-alternative forced-choice identification experiment, where the Centre of Gravity (CoG) was manipulated into seven steps across continua, the current study uncovered a significant reliance of Taiwan Mandarin listeners on lexicality when encountering ambiguous sibilant signals. The observed overall bias against the gap sibilant end aligns with the well-established lexicality effect on phonemic categorization. The present study extends this understanding by examining the lexicality effect within a context of ongoing sibilant merger, considering the complex interplay between frequency and gap factors in phoneme categorization. As mentioned earlier, in Taiwan Mandarin, the alveolar sibilants /s ts tsʰ/ and retroflex sibilants /ʂ tʂ tʂʰ/ undergo a process of merger primarily through the de-retroflexion of the retroflex category (Ing, 1984; Kubler, 1985). This merger leads to a reduction in the distance between the Centre of Gravity (CoG) of the two categories. Given the unidirectional nature of this merging process, it is likely that the alveolar sibilants are perceived as "de-retroflexed" variants of the retroflex sibilants, resulting in a less marked gap position on the continua. The prediction is confirmed by the results of a stronger effect size on the estimated coefficients where the alveolar end is presented as an acoustically ambiguous stimulus (i.e., GW), compared to its retroflex gap counterpart (i.e., WG).

Another novel finding in the current study is the expansion of the top-down categorization process interacting with ongoing merger progress. Participants in current study, different from the previous studies on lexicality effect (Ganong, 1980), frequency effect (Connine et al., 1993; Politzer-Ahles et al., 2020) or tone frequency effect (Yang et al., 2022); despite explicit informed about the presence of gap sounds and instructed to focus on the onset of the first syllable and provide rapid responses, results were still significantly influenced by the lexicality effect even at the endpoints of the continua. This finding suggests a strong inclination among participants to align their responses with the high-level linguistic context and prioritize the perception of complete words more heavily than previous studies. The findings of this study provide support for the nearly completed merger status of the onset sibilants in Taiwan Mandarin. The results indicate that the distinction between the alveolar and retroflex sibilant categories is generally indistinguishable in perception. This means that participants have difficulty perceiving a clear contrast between the two categories, even when presented with highly unambiguous acoustic signals. Additionally, the influence of lexicality on perception further emphasizes the impact of the merger, as it affects the perception of sibilant sounds regardless of their clarity. Alternatively, in word-word continua (WW) where lexicality effect is absent, a higher proportion of alveolar responses compared to retroflex responses is observed. Considering the randomized nature of all trials, it can be attributed to a frequency effect in the perception of alveolar sibilants.

The asymmetrical pattern between pairs remains. As previously mentioned, the z-zh [ts-tʂ] pair exhibits a notable resistance of retroflex responses in comparison to the other pairs in the gap-word (GW) continua. In contrast, the c-ch [tsʰ-tʂʰ] pair exhibits an opposite pattern, with an extreme lexical effect observed at the earliest step, resulting in more than half of the responses being retroflex at the alveolar gap endpoint (i.e. Figure 2, upper left plot). Following consideration of casual possibilities are considered: first, the variations of word frequency across all word pairs. (Connine et al., 1993) reported the influence of word frequency on phoneme identification tasks, showing a bias towards high-frequency words when it comes to ambiguous tokens. The issue was fully aware of and made significant efforts to carefully balance it. However, it is important to acknowledge the inherent unpredictability and complexity associated with word frequency, and caution must be exercised when interpreting the results. Determining the exact degree of an ongoing merger can be challenging. It has been argued by Chung (2006) that the variation in mergers is not categorical, but rather exists on a continuum with considerable range. Without certainty regarding the participants' individual merger status, the perception of sibilants is likely to be influenced by their own production, resulting in potential heterogeneity among the different pairs. While the current study does not focus specifically on mergers from different places of articulation, the observed variations among different pairs in the perceptual process offer valuable insights into the nature of perception bias. These findings emphasize the significance of considering these variations in future studies, as they can contribute to a better understanding of the complexities of phonetic perception and the influence of different phonetic factors.

Taken together, the results of the study contribute to our understanding of merger phoneme categorization in relation to gap and lexical effect.

5 Appendices

Wordlist		
Continua	Chinese	IPA
GW	查看	t̺ʂʰa kʰan
GW	扯淡	t̺ʂʰə tan
GW	舌頭	ʂɤ tʰou
GW	神經	ʂən t͡ɕiŋ
GW	這邊	t͡sə piən
GW	整修	t͡səŋ ɕiu
WG	彩虹	tsʰai hoŋ
WG	菜販	tsʰai pʰan
WG	鬆餅	soŋ peiŋ
WG	所以	suɔ i
WG	嘴巴	tsuɪ pa
WG	作品	t͡swɔ pʰĩ
WW	殘餘 - 單于	t͡sʰ/ t͡ʂʰ an y
WW	曾經 - 成精	t͡sʰ/ t͡ʂʰ ən t͡ɕiŋ
WW	三菱 - 山林	s/ʂ am liŋ
WW	桑葚 - 傷腎	s/ʂ aŋ ʂən
WW	髒話 - 彰化	t͡s/ t͡ʂ aŋ ua
WW	臟器 - 脹氣	t͡s/ t͡ʂ aŋ t͡ɕi

6 References

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using **lme4**. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Boersma, P., & Weenink, D. (1992, 2022). *Praat: Doing Phonetics by Computer*. https://www.praat.org.

Chung, K. S. (2006). Hypercorrection in Taiwan Mandarin. *Journal of Asian Pacific Communication*, *16*(2), 197–214. https://doi.org/10.1075/japc.16.2.04chu

Connine, C. M., Titone, D., & Wang, J. (1993). Auditory word recognition: Extrinsic and intrinsic effects of word frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*, 81–94. https://doi.org/10.1037/0278-7393.19.1.81

Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance*, *6*, 110–125. https://doi.org/10.1037/0096-1523.6.1.110

Ing, R. O. (1984). Issues on the pronunciations of Mandarin. *The World of Chinese Language*, *35*, 6–16.

Jongman, A., Wayland, R., & Wong, S. (2000). Acoustic characteristics of English fricatives. *The Journal of the Acoustical Society of America*, *108*(3), 1252–1263.

Kawahara, H., Morise, M., Takahashi, T., Nisimura, R., Irino, T., & Banno, H. (2008). Tandem-STRAIGHT: A temporally stable power spectral representation for periodic signals and applications to interference-free spectrum, F0, and aperiodicity estimation. *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*, 3933–3936. https://doi.org/10.1109/ICASSP.2008.4518514

Kubler, C. C. (1985). The Influence of Southern Min on the Mandarin of Taiwan. *Anthropological Linguistics*, *27*(2), 156–176.

Lee, C.-Y., Zhang, Y., & Li, X. (2014). Acoustic characteristics of Mandarin voiceless fricatives. *Journal of Chinese Linguistics*, *42*, 151–170.

Lee-Kim, S.-I., & Chou, Y.-C. I. (2022). Unmerging the sibilant merger among speakers of Taiwan Mandarin. *Laboratory Phonology*, *13*(1), Article 1. https://doi.org/10.16995/labphon.6446

Politzer-Ahles, S., Lee, K. K., & Shen, L. (2020). Ganong effects for frequency may not be robust. *The Journal of the Acoustical Society of America*, *147*(1), EL37–EL42. https://doi.org/10.1121/10.0000562

R Core Team (2020). —European Environment Agency. (n.d.). [Methodology Reference]. Retrieved May 21, 2023, from https://www.eea.europa.eu/data-and-maps/indicators/oxygen-consuming-substances-in-rivers/r-development-core-team-2006

Schneider, W., Eschman, A., & Zuccolotto, A. (2012). *E-Prime User's Guide. Pittsburgh: Psychology Software Tools, Inc.* https://support.pstnet.com/hc/en-us

Yang, T.-H., Jin, S.-J., & Lu, Y.-A. (2022). The Effect of Lexicality, Frequency, and Markedness on Mandarin Tonal Categorization. *Frontiers in Psychology*, *13*. https://www.frontiersin.org/articles/10.3389/fpsyg.2022.836865