SIBILANT PERCEPTION BY MERGED SPEAKERS: THE CASE OF TAIWAN MANDARIN

Sang-Im Lee-Kim & Hsiang-Yu Tung

HIPCS Hanyang University; National Yang Ming Chiao Tung University
sangimleekim@hanyang.ac.kr; loveh89301@gmail.com

ABSTRACT

In many sound changes, the merging of categories in perception often precedes the merger in production, while the reverse, namely merger in production but distinction in perception, has been argued to be rare [1]. The present study examined the perception of the alveolar-retroflex sibilants in Taiwan Mandarin (TM), focusing on the relationship between perception and production by merged speakers. The results of an AXB discrimination task showed that merged speakers were able to discriminate the sibilants above chance level but did so less accurately than non-merged distinct speakers. In an identification task with social guises, merged speakers were again shown to be less sensitive to phonetic cues and more sensitive to social cues. The results indicate that the frequent encounters with distinct forms in the speech community may have helped merged speakers keep the categories separate, to some extent, in their memory.

Keywords: sibilant merger, speech perception, Taiwan Mandarin, social guises

1. INTRODUCTION

In many sound changes, categories are often merged in perception before being merged in production [1]. This tendency translates to the frequent attestation of near-mergers—speakers maintain small but consistent distinctions between categories in production but fail to perceive the small differences in perception [2, 3]. The opposite pattern, namely merger in production but distinction in perception, has also been observed [4, 5, 6, 7, 8, 9]. For example, although speakers of the PIN-PEN merger were outperformed by their distinct counterparts, their identification accuracy was consistently higher than chance level [9]. In the case of the Cantonese tone merger, the perception of merged speakers was shown to remain largely intact in an AX discrimination, despite their slower responses than distinct speakers [10].

One possible reason for the apparent perception-production mismatch in phonemic mergers is that merged speakers are exposed to distinct speech in a speech community in which large variation exists among the speakers [9, 11]. Frequent encounters with the exemplars of those distinct tokens may keep the categories from being completely merged in their mental representation. It has also been proposed that the contrasts maintained in a non-merging context may help facilitate the perception of the distinct categories [8, 9]. For example, /ɪ/ and /ɛ/ are merged before nasals in the PIN-PEN merger, but the vowel contrast is robust in non-pre-nasal contexts. Merged speakers may, therefore, extend their sensitivity to the acoustic cues available in non-merging contexts to the merging context.

The present study examined the perception of the alveolar-retroflex sibilants in Taiwan Mandarin (TM), focusing on the relationship between perception and production by merged speakers. The sibilants in TM appear only in the syllable initial position and are known to be variably implemented, ranging from full merger to clear contrasts (e.g., /s~ʂ/) [12]. This pattern is unconditional as the merger is not limited to particular phonological contexts, and is clearly above the level of social awareness [13]. Two perception tasks (identification and discrimination), were conducted with two goals in mind. First, we aimed to rigorously examine the merged speakers' perception of the contrasts. It was pointed out that a commutation test may be too conservative as a measure to assess the merged speakers' actual perceptual ability [8, 9]. Often used in sociolinguistic research, commutation tests utilize 100% accuracy as the cut-off point; failure in this task is interpreted as the speaker being merged in perception. This measure is likely to underestimate the merged speakers' actual perceptual sensitivity [8], and a more sophisticated method was thus warranted.

Second, we investigated the combined effects of social expectations and one's production characteristics on the perception of the TM sibilants. The impact of socio-indexical information prevails in speech perception [14, 15, 16], and speakers of mergers-in-progress are also sensitive to the social background associated with particular phonetic variants [6]. The TM sibilant merger is socially structured in that it is often attributed to contact with a substratum language, Taiwanese Southern Min (TSM), which lacks the retroflex category in its inventory [17, 18]. We examined whether the model talker's social background, namely expected TSM fluency, would have a systematic influence on sibilant perception. The identification experiment was, therefore, designed to include socially motivated guises such that a model talker was described as being from either a city where TSM is not widely spoken or a city where TSM is more common.

2. EXP 1: SIBILANT IDENTIFICATION WITH SOCIAL GUISE

2.1. Participants

Seventy-three Taiwanese college students (40 females, 33 males, aged 20-29) participated in both production and perception experiments. A wordlist reading task was performed for all the participants to establish their merger status. The wordlist consisted of four disyllabic words with word-initial sibilants (alveolars /s ts^{h}/ vs. retroflexes /ʂ tʂ^{h}/) followed by the vowels /a/; many filler items were also included. The randomized wordlist presented in Chinese characters was repeated five times, and the speech signals were recorded at a sampling rate of 44.1 kHz.

The frication noise was labeled manually in Praat [19] and was submitted to a multitaper spectral analysis in Matlab [20] to obtain spectral mean (M1) and peak values. The mean values of the retroflexes were then subtracted from those of the alveolars for each individual speaker to obtain spectral peak and spectral mean distances, respectively. As shown in Figure 1, the two values were highly correlated ($r = .950$, $p < .0001$). Individuals' merger status was further determined through a two-sample *t*-test: a non-significant difference in the distribution of either spectral means or peaks was taken to indicate that the speaker was merged. Two groups were identified with respect to the merger status: 28 MERGED (11F, 18M) and 44 DISTINCT (29F, 15M) (Table 1). A linear regression model fitted to the spectral distance data in R [21] revealed a significant effect of sex but not TSM-fluency; males were more likely to merge the sibilants than females ($p = .0055$), but TSM-fluency ($p = .8486$) and its interaction with sex were not significant predictors of the merger ($p = .7234$). This is consistent with previous findings [12, 22] that merger is not necessarily triggered by TSM fluency—the pattern has become widespread throughout the speech community in Taiwan.

Table 1. Numbers of participants by merger status and sex.

	F	M	total
MERGED	11	18	29
DISTINCT	29	15	44
total	40	33	73

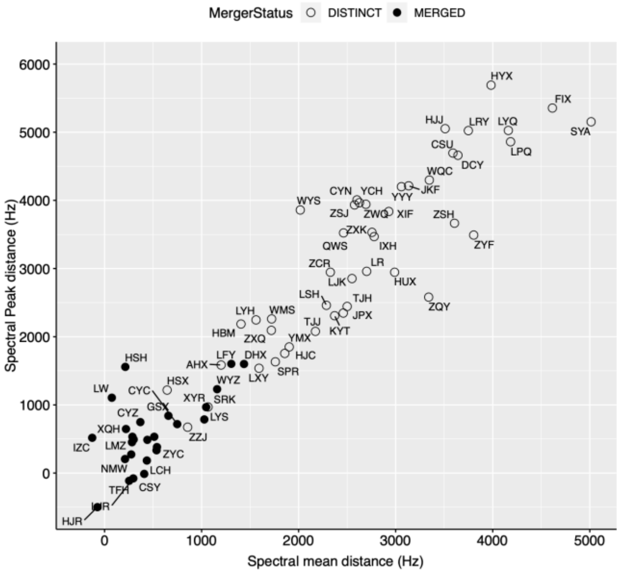

Figure 1. The distribution of spectral peak distance against spectral mean distance by each participant. MERGED speakers are marked with filled circles, and DISTINCT speakers with empty circles.

2.2. Stimuli

For the perception study, a TM male speaker who conveyed clear alveolar-retroflex distinction produced twelve monosyllabic words containing initial sibilants /s ts ts^{h}/ and /ʂ tʂ tʂ^{h}/ followed by /a/ and /u/ carried by Tone 1 (X^{55}). These original tokens were digitally manipulated in Praat. The frication noise /s/ and /ʂ/, for example, was first equated for duration and combined with different amplitude proportions to create 8-step continua. To increase acoustic ambiguity, digital manipulation was conducted twice (Figure 2). The vowels were spliced from the original alveolar tokens. Figure 3 presents the spectra of the /sa-ʂa/ and /su-ʂu/ continua. As shown in the figures, the spectral peaks at higher frequencies (8-9 kHz) gradually decreased in amplitude while those at lower frequencies (1-3 kHz) were amplified as the signals changed from most /s/-like to most /ʂ/-like tokens. Note that the acoustic difference of the sibilants from one end to the other was larger for the sibilants before /u/ than before /a/. This was not intended, nor expected, but would likely lead to perceptual consequences, which will be discussed later.

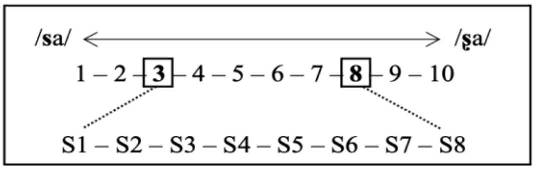

Figure 2. Acoustic manipulation for sibilant continuum

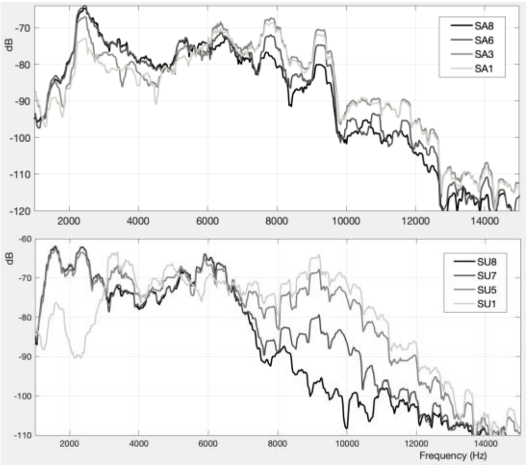

Figure 3. /s/-/ʂ/ spectral continua in the /a/ (top) and /u/ vowel context (bottom).

2.3. Procedure

The perception experiment was conducted in a sound-attenuated booth at National Yang Ming Chiao Tung University in Taiwan. Participants were randomly assigned to one of the social conditions: in the TAIPEI condition, they were told that the talker was originally from Taipei in Northern Taiwan (more likely to speak "standard" distinct variety TM), and in the TAINAN condition, they were told the talker was from Tainan in Southern Taiwan (more likely to speak TSM). To provide further social information, participants watched a video clip of a singer performing a popular song in their respective languages prior to the experiment, and landmarks of the respective cities were presented on the computer screen during the experiment (Figure 4).

Figure 4. Photos of representative buildings in each city: Anping Castle in Tainan (left) and the Taipei 101 in Taipei (right).

In a 2-AFC identification task, participants categorized the target stimuli as alveolar or retroflex by selecting the corresponding letter in the *Zhuyin* phonetic alphabet. There were three blocks in total, each of which consisted of 90 trials containing 48 targets (3 sibilant sets * 8 steps * 2 vowels) and 42 filler items. Fillers included various non-sibilant sounds (e.g., vowels [i] or [u]). The experiment was administered in E-prime 3 [22] and took about 20 minutes to complete.

2.4. Results

The results of the identification task are summarized in Figure 5. The data were analyzed using mixed-effects logistic regression in R. The model revealed three prominent findings. First, in the baseline condition (MERGERSTATUS = MERGED, VOWEL = /a/), merged speakers perceived more retroflexes in the TAIPEI than in the TAINAN condition ($p < .01$), reflecting their implicit bias that a speaker from Tainan would not produce "proper" retroflexes. This was also the case for the distinct speakers, as indicated by the lack of a significant interaction between MERGERSTATUS and SOCIALCONDITION. It is interesting that the speakers' biases were reflected in their perception, even though TSM fluency has lost predictive power on the sibilant merger in production (see §2.1). Second, the merged speakers were less sensitive to the fine-grained acoustic details of the stimuli than the distinct speakers (STEP*MERGERSTATUS: $p < .0001$). As signified by the flatter slopes, merged speakers were reluctant to make either choice even for the stimuli from the two ends, potentially leading to a greater reliance on social cues. Third, the effect of social cues was greatly attenuated in the /u/ vowel context, particularly for the merged speakers (SOCCOND*VOWEL: $p < .0001$). This finding seems to reflect the relative saliency of the acoustic cues in frication noise conditioned by the following vowels. Recall that the acoustic differences in frication noise were larger for the /u/ than for the /a/ vowel context (Figure 3). Cue-trading between social and phonetic cues is evident in this case. Taken together, the results showed that the merged speakers were less sensitive to phonetic cues but more sensitive to social cues.

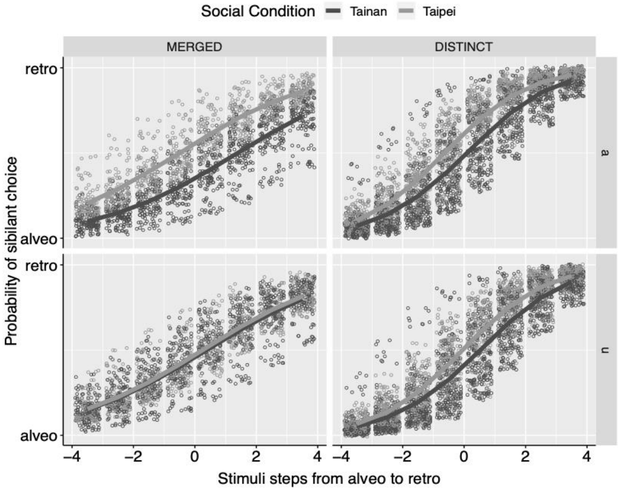

Figure 5. Predicted sibilant choices by the social guise broken down by the speakers' merger status and vowel context.

3. EXP 2: SIBILANT DISCRIMINATION

3.1. Participants

Thirty-five speakers were invited to participate in the discrimination task. All but two participants had participated in the identification task about six months before. Sixteen participants were from the MERGED group (4F, 12M) and nineteen were from the DISTINCT group (16F, 3M).

3.2. Stimuli and procedure

Among the tokens used for the identification task (Figure 2), S1 and S2 tokens were chosen as exemplars of alveolars and S7 and S8 tokens were as exemplars of retroflexes for the discrimination task. The two tokens from one end were similar but not identical acoustically, which was thought to make the experimental task sufficiently difficult. For example, the sequence S1(alveo)-S7(retro)-S8(retro) was created for an AXB trial, in which case the correct answer would be B. 96 tokens (4 pairs * 4 orders * 3 manners of sibilants * 2 vowels) were repeated three times across three blocks. Participants were given 3 seconds to respond in E-prime. The experiments were conducted individually in a sound-attenuated booth.

3.3. Results

The results of the discrimination task are summarized in Figure 6. The accuracy data were analyzed using a mixed-effects logistic regression, and the log response time data were analyzed using a mixed-effects regression in R.

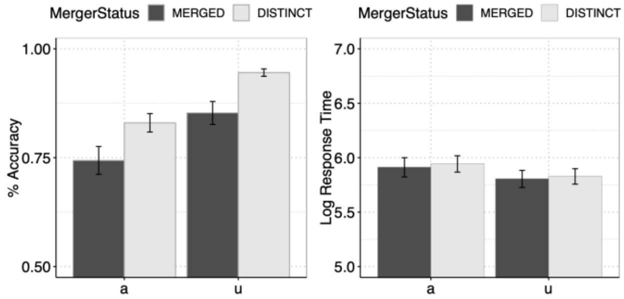

Figure 6. Mean discrimination accuracy (left) and log response time (right) by vowel and the speakers' merger status.

Two major findings were drawn from the model. First, the merged speakers' discrimination of the sibilants (M = 80%) was significantly less accurate than that of the distinct speakers (M = 88%) ($p < .01$), but they performed well above chance level, in line with previous works [4, 5, 6, 7, 8, 9]. Second, both the discrimination accuracy and response time data converged on the vowel effect: the sibilants were discriminated more accurately ($p < .0001$) and faster ($p < .0001$) in the /u/ than in the /a/ context, regardless of the speaker's merger status. Consistent with the results of the identification task, this can be attributed to the larger acoustic differences in the noise properties before /u/ than before /a/.

4. DISCUSSION

This study investigated the perception of the TM alveolar-retroflex sibilants by merged speakers. The results showed that merged speakers were able to discriminate the sibilants above chance level, but did so less accurately than distinct speakers. The TM sibilant merger is unconditional, and merged speakers' moderate perceptual ability was thus likely to arise from their exposure to distinct forms carried by other speakers in the speech community. In addition, the merged speakers' sensitivity to the distinct categories is likely to be further enhanced by the high social awareness of this variant. The sibilant merger is subject to overt comments among TM speakers [13], which may have facilitated the maintenance of the separate exemplars for the two categories. The results of the identification task accompanied by the social guises lend further support for this analysis. Merged as well as distinct speakers in this task were sensitive to the social cues, in which the implicit biases systematically modulated the perception of the sibilants. This result demonstrates the pattern—merger in production and distinction in perception in phonemic mergers—may be more widely attested than previously assumed and should, therefore, be reconsidered as a genuine linguistic pattern of sound change.

The vowel context effects are also worth discussing. In the identification task, merged speakers, in particular, showed a greater weighting of phonetic cues in the /u/ vowel context accompanied by reduced reliance on social cues; in the discrimination study, both groups performed better in the /u/ than in the /a/ vowel context. The acoustic differences in the frication noise were greater in the rounded vowel context (Figure 3), which seems to have reduced the reliance on non-acoustic cues in speech perception. Lip rounding of the following vowel may enhance retroflexion of the retroflex category [24], which could explain the salient acoustic cues in the /u/ context and hence the reduced reliance on other cues.

5. ACKNOWLEDGMENTS

We would like to thank Yun-Chieh Chou, Kuei-Hong Lin, Wei-Cheng Weng for help with data collection and processing. Special thanks also to Yu-an Lu, Li-Hsin Ning, and Meredith Tamminga for their helpful comments. This work was supported by MOST110-2410-H-A49-058-MY3, Taiwan and start-up research funds from Hanyang University, Korea.

6. REFERENCES

[1] Labov, W. 2011. *Principles of Linguistic Change. Vol. 3: Cognitive and Cultural Factors*. John Wiley & Sons.

[2] Labov, W., Yaeger, M., Steiner, R. 1972. A quantitative study of sound change in progress. In *U.S. Regional Survey*. Philadelphia.

[3] Yu, A. C. L. 2007. Understanding near mergers: the case of morphological tone in Cantonese. *Phonol* 24, 187–214.

[4] Hay, J., Drager, K., Thomas, B. 2013. Using nonsense words to investigate vowel merger. *Eng. Lang. Ling.* 172, 241–269.

[5] Thomas, B., Hay, J. 2005. A pleasant malady: The ellen/allan merger in New Zealand English. *Te Reo* 48.

[6] Hay, J., Warren, P., Drager, K. 2006. Factors influencing speech perception in the context of a merger-in-progress. *J. Phon.* 344, 458–484.

[7] Baranowski, M. 2013. On the role of social factors in the loss of phonemic distinctions. *Eng. Lang. Ling.* 172, 271–295.

[8] Wade, L. 2017. The role of duration in the perception of vowel merger. *Lab. Phon. 8*(1), 1–34.

[9] Austen, M. 2020. Production and perception of the Pin-Pen merger. *J. Ling. Geog.* 82, 115–126.

[10] Mok, P. P. K., Zuo, D., Wong, P. W. Y. 2013. Production and perception of a sound change in progress: Tone merging in Hong Kong Cantonese. *Lang. Var. Chan.* 25, 341–370.

[11] Maguire, W., Clark, L., Watson, K. 2013. Introduction: what are mergers and can they be reversed? *Eng. Lang. Ling.* 17(2), 229–239.

[12] Lee-Kim, S., Chou, Y. I. 2022. Unmerging the sibilant merger among speakers of Taiwan Mandarin. *Lab. Phon.* 13.

[13] Chung, K. S. 2006. Hypercorrection in Taiwan Mandarin. *J. As. Pac. Comm.* 16(2), 197–214.

[14] Niedzielski, N. 1999. The effect of social information on the perception of sociolinguistic variables. *J. Lang. Soc. Psy.* 181, 62–85.

[15] Hay, J., Drager, K. 2010. Stuffed toys and speech perception. *Ling.* 48(4), 865–892.

[16] Drager, K. 2010. Sociophonetic variation in speech perception. *Lang. Ling. Comp.* 47, 473–480.

[17] Ing, R. 1984. Issues on the pronunciations of Mandarin. *Wor. Chn. Lang.* 35, 6–16.

[18] Kubler, C. C. 1985. The influence of Southern Min on the Mandarin of Taiwan. *Anth. Ling.* 272, 156–176.

[19] Boersma, P., Weenink, D. 2020. Praat: doing phonetics by computer [Computer program]. Version 6.0. 37. 2018.

[20] Blacklock, O. S., Shadle, C. H. 2003. Spectral moments and alternative methods of characterizing fricatives. *J. Acoust. Soc. Am.* 113(4), 2199.

[21] R Core Team. 2022. *R: A language and environment for statistical computing*.

[22] Chuang, Y.-Y., Sun, C.-C., Fon, J., Baayen, R. H. 2019. Geographical variation of the merging between dental and retroflex sibilants in Taiwan Mandarin. *Proc. 19th ICPhS.* Melbourne.

[23] Schneider, W., Eschman, A., Zuccolotto, A. 2002. *E-Prime: User's guide*. Psychology Software Incorporated.

[24] Rau, D., Li, J. 1994. Phonological variation of (ts), (tsh), and (s) in Mandarin Chinese. *Proc. 23rd NWAV*, Stanford University.