Sibilant production in Taiwan Mandarin: untangling the effects of linguistic and social variables

Yu-Leng Lin [1,*], Jiun-Shiung Wu[2]

1,2 National Chung Cheng University, Institute of Linguistics, Chiayi, Taiwan
* Correspondence: Lngyllin@ccu.edu.tw

Abstract

This study examines dental and retroflex sibilants in Taiwan Mandarin. The center of gravity of frication and formant transitions of the following vowel are used to examine whether a dental-retroflex merger is present in Taiwan Mandarin. The factors that condition this merger are also examined, including region and gender. The results indicate that contrasts between dentals and retroflexes were, in general, preserved across regions and genders. In addition to acoustic analysis, the results of machine learning analysis indicate that center of gravity, F2 and Min proficiency were effective indexes for distinguishing dentals from retroflexes.

1 Introduction

Taiwan Mandarin is usually described as distinguishing retroflex from dental fricatives/affricates. However, as previous studies have reported, this distinction is not present in some varieties of Taiwan Mandarin (henceforth, Mandarin). In these varieties, retroflexes and dentals are merged in favor of dentals, and the presence of this merger has been reported to be correlated with social factors such as gender, region of origin and age (see Chuang, 2009; J. M.-C. Li, 1995; Lin, 1983; Tse, 1998 for more discussion). This merger, or deretroflexion, has been ascribed to the influence of Taiwan Southern Min (henceforth, Min), a language that lacks contrastive retroflexion (Kubler, 1985). This near or complete merger has been claimed to be one of the most prominent differences between Mandarin as spoken in the south of Taiwan, where it is more prevalent (especially among male speakers), and the north of Taiwan, where it is less prevalent (Chuang, 2009; J. M.-C. Li, 1995; Lin, 1983).

Level of Min proficiency seems a good candidate for explaining the variable degree of merger in Mandarin-Min speakers: under this explanation, the lack of contrastive retroflexion in Min leads to a merger of retroflexes and dentals by means of deretroflexion when Mandarin-Min bilinguals speak Mandarin (Lin, 1983). Min interference had, indeed, been the dominant explanation in the literature (e.g., Ing, 1984; Kubler, 1985) until the late 2010s. However, the most recent work on the subject has shown no effect of Min proficiency on merger (Chuang, Sun, Fon, & Baayen, 2020; Kuo, 2018; Lee-Kim & Chou, 2022; Lu, 2019). How are we to reconcile these results?

This paper argues that the methods of assessing Min proficiency in the recent literature lack ecological validity: these studies used self-reported proficiency, which is known to be unreliable (Blue, 1994; Shameem, 1998). When we use a more reliable measure, namely third-party rating using native speaker judges, the effect of Min proficiency re-emerges. We have also found, contrary to Chuang et al. (2020), Kuo (2018) and Lu (2019), that contrasts between dentals and retroflexes were, in general, preserved across regions and genders. If contrasts between dentals and retroflexes are preserved rather than lost, to what extent does each of the factors of interest contribute to the distinction between the two sets of sounds? To examine this, we subjected the results to a machine learning analysis, and we found that center of gravity (COG) of frication was the most effective index for distinguishing dentals from retroflexes. Adding either Min proficiency or F2 to COG in the model helped raise the model's accuracy in distinguishing dentals from retroflexes.

The paper is organized as follows. Section 2 presents the method. Section 3 presents and discusses the results. Section 4 concludes with a general discussion and draws implications for future research.

2 Method

2.1 Participants

Twenty-nine participants were recruited in Toronto, Canada. The acceptability criteria for the study were as follows: (1) participants recruited had to be between 18–40 years of age at the time (born between 1995 and 1974), (2) they needed to be speakers of Taiwan Mandarin and Min, (3) they must have lived in the same city or county until 13 years of age and (4) they needed to have normal speech, hearing, and vision. Participants were recruited from the population of Taiwanese graduate students studying at the University of Toronto (via a post on the Facebook group Taiwanese Graduate Student Association in Toronto (TGST)) and from Taiwanese people present in Toronto on a working holiday visa (via a post on the Facebook group Working Holiday in Canada - Toronto).

Of the male participants, six were from Northern Taiwan (defined as one of the cities/counties of Taipei, Taoyuan, Kinlong, Hsintzu, or Miaoli) and eight were from Southern Taiwan (defined as one of the cities/counties of Chiayi, Tainan, Kaohsiung, or Pingtung). Of the female participants, eight were from Northern Taiwan and seven were from Southern Taiwan. Twenty-nine of the participants reported that their parents spoke and used Min, and two reported their parents did not speak Min but that they were exposed to Min and conversed in Min with friends. Of these two, one was a northern male and the other was a northern female. All spoke L2 English, which they started learning at approximately age 12. All had between 3 to 9 years' experience living in an English-speaking country. None was linguistically trained.

There is a well-established effect of age on the presence of the dental-retroflex merger in the literature (Kubler, 1985; Lin, 1983). As our focus was on gender and region of origin, however, we limited the participant pool to members of the generation who were between 18–40 years of age, so that we could ensure a sufficient sample size in each group for the variables of interest. Extending this analysis to encompass variation across age groups is a logical next step in pursuing this research question.

2.2 Min proficiency test

To examine Min influence, a Min proficiency test was given to participants to examine how proficiently they comprehended and produced Min. In this Min test, participants heard one word at a time in Min, and then they were asked to explain/describe what it referred to. They could answer either in Mandarin or Min, with responses in Min encouraged. The words for this test were chosen from Chiang (2011), which divides words into proficiency levels ranging from basic, intermediate, and proficient. This source was chosen because it has been recognized as one of the reference books used for the two official Min proficiency tests, the one held by National Cheng Kung University and the other by the Ministry of Education.

A five-point Likert scale was used by two raters who were native Min speakers, ranging from 1 (totally wrong or unanswered), 2 (participant vaguely knew the answer), 3 (partially correct), 4 (totally correct, if the participant answered in Mandarin) and 5 (totally correct, if the participant answered in Min). Although the target word was in Min and participants were encouraged to answer in Min, they could choose to respond in either Mandarin or Min. To ascertain that two raters' Min proficiency scores were reliable, inter-rater reliability was examined using a Pearson correlation test. The correlation score of the two raters was 0.91 ($t = 73.799$, $df = 1158$, $p < 0.001$), which was above the threshold of 0.7, and was therefore considered good.

2.3 Mandarin production task

Sample stimuli for the production experiment are given in Table 1. In the Mandarin production task,[1] the test materials were disyllabic words. The task followed a 2 × 3 × 4 design: 2 places of

[1] Participants were also asked to read a Mandarin text (the material was the North Wind story written in Mandarin). Participants read the story aloud, which was used as a reference to ascertain whether they could speak Mandarin fluently.

articulation (dental, retroflex), 3 manners (fricative, unaspirated affricate, aspirated affricate), and 4 following-vowel contexts (a, Ø, ou, u). The consonants consisted of the voiceless coronal fricatives and affricates (9 consonants), which were placed in the syllable-initial position of the first syllable and combined with nuclei [ia] vs. [a]; [i] vs. Ø (empty nucleus);[2] [iou] vs. [ou]; and [y] vs. [u] (36 test items in total). Note that in this study, following the assumption of Svantesson (1986), nuclear [ia] and [a] were treated as /a/, [i] and Ø as /Ø/, [iou] and [ou] as /ou/, and [y] and [u] as /u/ based on their allophonic distribution following the dental sibilants [ts], [tsʰ], [s] vs. the palatal sibilants [tɕ], [tɕʰ], [ɕ]. To recapitulate the pattern, palatal sibilants occur before front vowels (i.e., [i], [y]) or diphthongs whose first element is a front vowel (e.g., [ia], [iou]), while dental sibilants occur with non-front vowels—complementary distribution.[3] We also controlled for tonal values, such that the first character had either Tone 1 (**5**5) or Tone 4 (**5**1) (because the beginning of the tone contour for both Tone 1 and Tone 4 is high (5)). To avoid tongue twisters, we ensured that the onset of the second syllable was non-coronal. There were 24 test items in total.

To avoid hypercorrective production, pseudorandomization was adopted, where disyllabic fillers (real words) were interspersed with test materials. The fillers consisted of disyllabic words containing consonant onsets other than the sibilants under investigation ([s, ts, tsʰ, ʂ, tʂ, tʂʰ]) combined with the vowels [a, Ø, ou, u] in the first syllable (46 fillers in total). An attempt was made to ensure that, for all test and filler items, the tonal value for the first syllable was either first or fourth tone; however, three filler items did not obey this restriction. We also tried to control the onset of the second syllable such that it was non-sibilant, but one filler item did not obey this rule.

1st syllable consonant	1st syllable vowel	1st syllable tone	2nd syllable tone	Disyllabic word (segmental IPA)	English gloss
[ts]	[a]	1	1	[tsa] [kən]	to root
[tsʰ]	[a]	1	2	[tsʰa] [mən]	to scrub the door
[s]	[a]	1	3	[sa] [xuãŋ]	to lie
[tʂ]	[a]	4	4	[tʂa] [pʰiɛ̃n]	to defraud
[tʂʰ]	[a]	1	2	[tʂʰa] [piɛ]	differen ce
[ʂ]	[a]	1	1	[ʂa] [tʰãn]	sandy beach

Table 1: Sample target items included in the Mandarin production task.

2.4 Procedure

The Mandarin production experiment was run with E-Prime (Schneider, Eschman, & Zuccolotto, 2002). Participants were given oral instructions by the experimenter and read detailed written instructions on the computer screen. The experiment took place at the sound-booth lab at the University of Toronto.

A series of disyllabic words written using traditional Chinese characters appeared on the screen. Participants were instructed to pronounce each word naturally, as they would in their daily speech, rather than in standard Mandarin. They were asked to repeat each disyllabic word six times in a row, yielding 144 (24 × 6) test tokens and 276

[2] Some scholars have analyzed the empty nucleus as [ɨ] (see Svantesson, 1986; Tse, 1998 among others).

[3] Svantesson (1986, p. 55) argues that palatals result from an historical sound change from alveolars and velars in palatalizing environments.

[*s], [*x] > [ɕ]

[*ts], [*k] > [tɕ]

[*tsʰ], [*kʰ] > [tɕʰ]

(46 × 6) filler tokens per participant. Tokens were recorded using a recorder (Zoom H4n).

2.5 Acoustic analysis

One primary coder and one additional coder used Praat to transcribe the sound files collected from the Mandarin production task and annotate the boundaries of frication and vowels. The values of measurements of the annotated files were extracted by a custom-written Praat script (Boersma & Weenink, 2021; version 6.1.53). Three measurements were taken: (1) one spectral moment of frication noise, (2) the COG and (3) F2. The data were recorded at a 44.1 kHz sampling rate. For spectral measurements, the onset of frication was annotated at the first appearance of white noise in a frequency band above 1000 Hz in the spectrogram and the end of the frication was annotated at the start of the following vowel (F. Li, 2008). A 23.2 ms Hamming window from the middle of frication was used for spectral analysis. Prior to the spectral analysis, all sound was modified to filter out frequencies below 500 Hz (pass Hann band filter with a lower limit of 500 Hz and upper limit of 22050 Hz in Praat).

When measuring F2 for female participants, the LPC (linear predictive coding) formant ceiling settings were set to 5500 kHz; for male participants, the formant ceiling was set to 5000 kHz. For females, the number of formants was set to five for all vowel contexts; for males, the number of formants was set to six for back vowels and five for other vowels. The formants of the following vowels were measured from an onset of 0.0125 seconds.

Only the target segments annotated by the primary coder were used for statistical analyses. The purpose of using an additional coder was to ascertain that the primary coder's annotated data were objective. Inter-rater reliability was examined using a Pearson correlation test. The Pearson product-moment correlations showed a correlation between the primary and additional coder of 0.99 for COG (t = 212.62, df = 272, p < 0.001) and 0.99 for F2 (t = 416.42, df = 272, p < 0.001), which was above the threshold of 0.7, and was therefore considered good.

2.6 Machine learning analysis

Feature engineering/extraction is a machine learning technique for exploring which features function better for a classification task among a set of many features. In this paper, we follow the idea presented in Ozdemir and Susarla (2018) to examine which feature or features (i.e., factors of interest, namely COG, F2, Min proficiency, region, gender in our study) are best able to distinguish dental from retroflex consonants.

Essentially, we explore what feature or features perform best in predicting whether a consonant is dental or retroflex. In order to automatically determine the performance of different sets of features in predicting the distinction between the two places of articulation, we adopted the module sklearn.model_selection.GridSearchCV from the machine learning toolkit *scikit-learn* (Pedregosa et al., 2011). GridSearchCV takes as its parameters a model, a set of parameters for the model, the training data and the corresponding labels of the training data. It returns the best parameters and the accuracy obtained using the best parameters.

Moreover, following Ozdemir and Susarla (2018), we use four statistic-probabilistic classifiers: logistic regression, *k*-Nearest Neighbors, Decision Tree and Random Forest. A very brief explanation of these four classifiers is given below.

A logistic regression classifier tries to draw a line that can separate the data as best it can. Given a data point, *k*-Nearest Neighbors finds the data points closest to that data point and assigns the point to the category to which most of the nearest data points belong.

Decision trees' aim is to craft a model for the prediction of a target variable's value. This is achieved by the construction of decision rules, which are produced by inference from the features of the data (Pedregosa et al., 2011).

Random Forest is an algorithm that combines several decision trees to produce a single result.

In this study, we used these four classifiers to examine how a specific set of features can predict the distinction between dentals and retroflexes. For any given feature, we try to see how well it can predict whether a sound is a dental or a retroflex. The higher the accuracy is when using a given feature, the larger role we conclude that this feature plays in distinguishing between these places of articulation.

3 Results and discussion

3.1 Acoustic analysis

Participants' Mandarin production data for the test items containing retroflex ([ʂ, tʂ, tʂʰ]) and dental ([s, ts, tsʰ]) sibilants were analyzed. Altogether, 144 tokens per participant were analyzed, yielding 4176 tokens in total. Mixed-effect regression modeling was used, as implemented in the R package lme4 (Bates, Mächler, Bolker, & Walker, 2015) in R version 4.1.1 (R Core Team, 2021). Since this study had two dependent variables, COG and F2, two models, containing one dependent variable each, were conducted separately. Note that the threshold for significance was set to $p < 0.025$ (= 0.05/2) due to the presence of two hypotheses (COG and F2). Both COG and F2 values were normalized and converted into z-scores.

The Independent variables were *place* (dentals vs. retroflexes), *region* (Southern Taiwan vs. Northern Taiwan), *gender* (male vs. female), *Min proficiency* (Likert scale 1–5) and *vowel context* (/a, Ø, ou, u/). For *place*, 'retroflex' was set as the reference group. In the COG model, given that the place of articulation for dental articulation is more fronted than for retroflex articulation, significantly higher COG values would be expected for dentals than for retroflexes. By this logic, if two-way contrasts were preserved, a positive effect of *place* would be expected in the COG model. However, for the F2 model, given that F2 values for the places of articulation are arranged as follows: palatal > retroflex > dental (Chiu, 2009), a negative effect of *place* would be expected if the distinction were preserved.

For *region*, 'Southern Taiwan' was set as the reference group. If participants from the north preserved two-way contrasts better than participants from the south, as predicted by the previous literature (J. M.-C. Li, 1995; Tse, 1998), a positive effect of *region* would be expected. For *gender*, 'male' was set as the reference group. If females were more likely to preserve two-way contrasts than males, as has been shown in the previous literature (Chuang, 2009; J. M.-C. Li, 1995; Lin, 1983), a positive effect of *gender* would be expected. If the degree of merger of dentals and retroflexes differed by region, an interaction between *place* and *region* would be expected. If the degree of merger were indeed the highest for southern males because of an interaction between region and gender, a three-way interaction between *place*, *region* and *gender* would be expected.

If *Min proficiency* were correlated with the merger of dentals and retroflexes, as indicated by Kubler (1985), an interaction between *Min proficiency* and *place* would be expected. If participants showed context-dependent overlap of dentals and retroflexes in vowel contexts, as reported in Chiu, Wei, Noguchi, and Yamane (2019), an interaction between *place* and *vowel context* would be expected.

The results for the COG model were as follows: dentals had a mean COG of 7963.72 (*sd* = 1854.59), while retroflexes had a mean COG of 5612.97 (*sd* = 1344.30). The mean COG for the various vowel contexts were 6001.88 for /a/ (*sd* = 1140.95), 7201.88 for /Ø/ (*sd* = 1530.21), 6528.86 for /ou/ (*sd* = 1931.46) and 6271.50 for /u/ (*sd* = 2026.86). The results for the F2 model were as follows: dentals had a mean F2 of 1424.83 (*sd* = 344.82), while retroflexes had a mean F2 of 1466.82 (*sd* = 351.78). The mean F2 values for the various vowel contexts were 1941.87 for /a/ (*sd* = 339.02), 2164.75 for /Ø/ (*sd* = 388.88), 1257.80 for /ou/ (*sd* = 214.75) and 1206.75 for /u/ (*sd* = 251.24). A scatterplot of dentals and retroflexes by COG and F2 is presented in Figure 1, along with an ellipse assuming a multivariate *t*-distribution at the confidence level of 0.95.

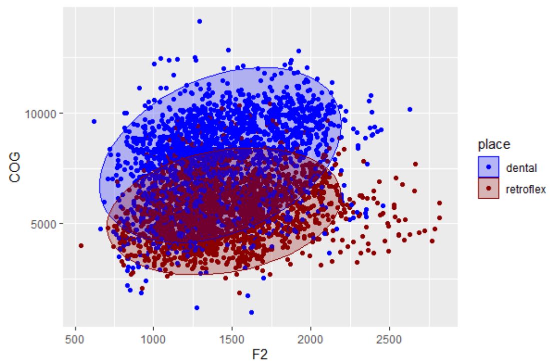

Figure 1. Dentals and retroflexes by COG and F2

The results for the COG model were as follows: The model found a positive effect of *place* ($p < 0.001$), suggesting that participants preserved two-way contrasts such that the place of articulation of dentals was significantly more fronted than that of retroflexes. There was a significant main effect of *vowel context* ($p < 0.001$) but no interaction between *vowel context* and *place* was found ($p > 0.025$), suggesting that there was no context-dependent overlap, at least with respect to COG values.

No main effect of *gender* ($p > 0.025$) or *region* ($p > 0.025$) was found. However, a main effect of *place* was positively significant ($p < 0.001$) along with a significant positive interaction between *place* and *region* ($p < 0.001$), as shown in Figure 2.

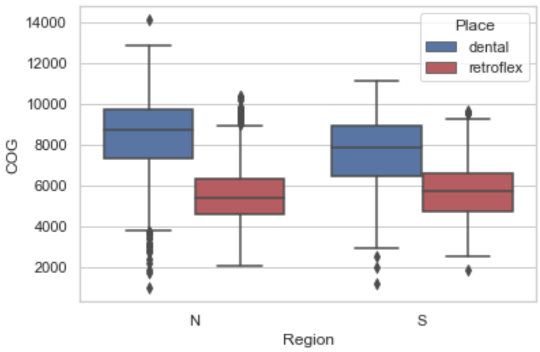

Figure 2. Interaction between place and region (COG) (N: North, S: South)

Although the literature has mentioned that Min proficiency tends to be higher in the south than in the north, no interaction effect between region and Min proficiency was found ($p > 0.025$). However, the interaction between *place* and *Min proficiency* was found to be positively significant ($p < 0.001$) (see Figure 3).

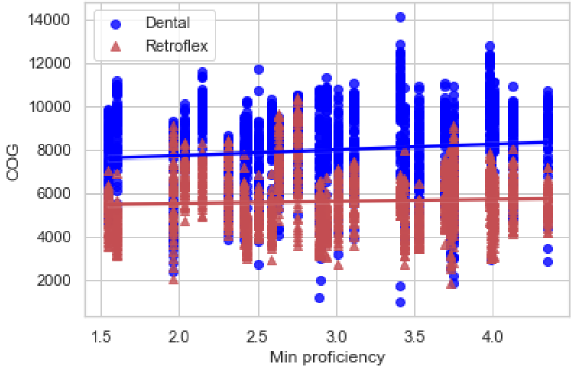

Figure 3. Interaction between place and Min proficiency (COG)

The results of the F2 model showed a slightly different situation. Contrary to the findings of the COG model, there was no main effect of *place* found ($p > 0.05$), but the interaction between *place* and *gender* was found to be positively significant ($p < 0.01$). A main effect of *gender* was found to be positively significant ($p < 0.001$), again unlike the COG model. The results suggest that dentals and retroflexes did not differ from each other significantly in F2 values in general. However, F2 values did differ between males and females, as shown in Figure 4. As was the case for the COG model, however, no significant main effect of *region* ($p > 0.05$) was found for the F2 model. Like the COG model, the F2 model found a significant main effect of *vowel context* ($p < 0.001$) but no interaction between vowel context and place, indicating that there was no context-dependent overlap for both COG and F2 values. As was the case for COG, however, no significant main effect of *region* ($p > 0.05$) was found for the F2 model. Like the COG model, the F2 model found a significant main effect of *vowel context* ($p < 0.001$) but no interaction between vowel context and place, indicating that there was no context-dependent overlap for both COG and F2 values. As was the case for COG, however, no significant main effect of *region* ($p > 0.05$) was found for the F2 model.

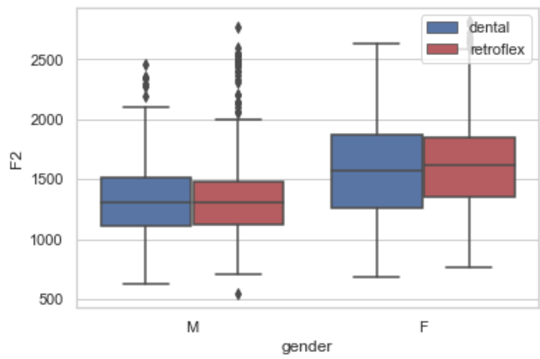

Figure 4. Interaction between place and gender (F2)

In contrast to Chuang (2009) and Tse (1998), who found that region and gender were the factors most strongly predicting sibilant merger, this study's results for these sociodemographic characteristics were more modest. Specifically, this study's MANOVA results showed that a complete merger of dental and retroflex sibilants occurred only for two of the southern male participants. This was similar to Tse's (1998) finding that female speakers preserved contrasts of dentals and retroflexes, while male Min speakers showed no distinction between the two, though the degree of merger found in the current study seems to be weaker than has been found in previous studies.

Taking COG and F2 together, all four groups (southern males/females, northern females/males) showed contrasts between dentals and retroflexes (as indicated by MANOVA), suggesting that contrasts between dentals and retroflexes were, in general, preserved across regions and genders.

3.2 Machine learning analysis

To determine what factors of interest (i.e., COG, F2, Min proficiency, region, gender) contribute to the distinction between dental and retroflex contrasts in natural language processing, four classifiers were adopted, namely, logistic regression, *k*-Nearest Neighbors, Decision Tree and Random Forest. Center of gravity (COG) of frication was found to be the most effective index for distinguishing dentals from retroflexes in all four classifiers (accuracy of distinction: 77.8%, 75.7%, 77.2%; 77.5%).

None of F2, Min proficiency, region or gender alone was found to be an effective index (accuracy of distinction ranging from 49.4% to 53.1% for the four classifiers).

F2 did, however, slightly facilitate the accuracy of distinction when it was combined with COG as a predictor of dental and retroflex contrasts in three of the four classifiers (accuracy of distinction: **79.6%**, **75.9%**, 77.2%, **78.3%**).

Min proficiency also slightly facilitated the accuracy of distinction when it was combined with COG as a predictor of dental and retroflex contrasts in three of the four classifiers (accuracy of distinction: 77.6%, **76.5%**, **77.7%**, **78.7%**).

4 Discussion and implications

This study has examined the effect of region, gender and Min proficiency on the merger of retroflex and dental coronal fricatives and affricates in Taiwan Mandarin, and has furnished a way to see what factors facilitate the accuracy of distinguishing between dentals and retroflexes, with two goals: (a) to examine if the merger of dentals and retroflexes can still be observed in the current populations in terms of COG and F2 and (b) to determine what factors contribute to the contrast between dentals and retroflexes.

Unlike previous work on the topic, which used self-reported Min proficiency, this study used third-party native speaker rating of participants' Min proficiency. Third-party rating using a Likert scale was chosen rather than self-rating because research has shown that the self-ratings reported by participants were not reliable. For example, as Blue (1994) has reported, learners of a language report unrealistically high or low levels of language proficiency when compared to the rating given to their proficiency by the instructors who were teaching them the language. Self-reported language ability has also been found to be inaccurate under conditions where the target language is perceived as lower in status than the dominant language of the community (Shameem, 1998).

To restate our methodology briefly, we used third-party native speakers to judge participants' Min proficiency based on whether they knew a set list of vocabulary items and whether they could explain their meaning fluently in Min. When Min proficiency was measured in this way, its effect on sibilant merger (as measured by COG and F2) became apparent. In addition, we believe that using third-party native speaker rating might better reflect the nature of Min proficiency in the population, which exists on a continuum.

Overall, the current study showed that contrasts between dentals and retroflexes were for the most part preserved across regions and genders. It also showed that COG is the primary feature distinguishing dentals from retroflexes. Finally, we have shown that F2 and Min proficiency still play a role in facilitating the distinction of dentals from retroflexes, even though the facilitation was not as strong as that provided by COG.

Acknowledgements

Thanks for Yoonjung Kang's, Alexei Kochetov's, Keren Rice's, and Jessamyn Schertz's comments and thanks to the SIG grant for covering a portion of the participants' fees.

References

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1-48.

Blue, G. M. (1994). Self-Assessment of Foreign Language Skills: Does It Work? *CLE Working papers, 3*, 18-35.

Boersma, P., & Weenink, D. (2021). Praat: doing phonetics by computer [Computer program].

Chiang, W.-w., Liang-wei Robert Cheng, Yu-hung Chang, Yuangshan Chuang, Ún-giân Iû°, Hung-yu Chang. (2011). *GTPT Taiwanese World Lists in Accordance with CEFR (全民台語認證語詞分級寶*

典). National Cheng Kung University: Center for Taiwanese Languages Testing.

Chiu, C. (2009). Acoustic and auditory comparisons of Polish and Taiwanese Mandarin sibilants. *Canadian Acoustics, 37*(3), 142-143.

Chiu, C., Wei, P.-C., Noguchi, M., & Yamane, N. (2019). Sibilant fricative merging in Taiwan Mandarin: An investigation of tongue postures using ultrasound imaging. *Language and speech, 63*(4), 877-897.

Chuang, Y.-Y. (2009). An Acoustic Study on Voiceless Retroflex and Dental Sibilants in Taiwan Mandarin Spontaneous Speech. *Taipei: National Taiwan University MA thesis*.

Chuang, Y.-Y., Sun, C.-C., Fon, J., & Baayen, R. H. (2020). *Geographical variation of the merging between dental and retroflex sibilants in Taiwan Mandarin*. Paper presented at the The 19th International Congress of Phonetic Sciences(ICPhS), Melbourne, Australia.

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

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

Kuo, J. (2018). A large-scale smartphone-based sociophonetic study of Taiwan Mandarin. *Asia-Pacific Language Variation, 4*(2), 197-230.

Lee-Kim, S.-I., & Chou, Y.-C. (2022). Unmerging the sibilant merger among speakers of Taiwan Mandarin. *Laboratory Phonology, 13*(1).

Li, F. (2008). *The phonetic development of voiceless sibilant fricatives in English, Japanese and Mandarin Chinese.* The Ohio State University,

Li, J. M.-C. (1995). *A sociolinguistic variation study of Chinese retroflex initials /tʂ/, /tʂʰ/, and /ʂ/ in Taiwan.* MA Thesis, Providence University, Taichung.

Lin, C. C. (1983). A sociolinguistic study of the use of the retroflex sounds in Mandarin in College students in Taiwan. *National Central University College of Arts and Letter Bulletin, 1*, 1-15.

Lu, W.-R. (2019). *Sibilant merger by Taiwanese-Mandarin bilingual speakers: a word processing and production study.* (MA). National Chiao Tung University,

Ozdemir, S., & Susarla, D. (2018). *Feature Engineering Made Easy: Identify unique features from your dataset in order to build powerful machine learning systems*: Packt Publishing Ltd.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., . . . Dubourg, V. (2011). Scikit-learn: Machine learning in Python. *The Journal of Machine Learning Research, 12*, 2825-2830.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime reference guide*: Psychology Software Tools, Incorporated.

Shameem, N. (1998). Validating self-reported language proficiency by testing performance in an immigrant community: The Wellington Indo-Fijians. *Language Testing, 15*(1), 86-108.

Svantesson, J.-O. (1986). Acoustic analysis of Chinese fricatives and affricates. *Journal of Chinese Linguistics*, 53-70.

Tse, J. K.-P. (1998). Do the younger generation in Taiwan really make no distinction between /ʐ, tʂʰ, tʂ, /, and /ts, tsʰ, s/ /? (臺灣地區年輕人ㄓㄔㄕ與ㄗㄘㄙ真的不分嗎？). *Mandarin world (華文世界)*, 1-7.