Production of the Taiwanese Southern Min tones by L1 Mandarin speakers

Karen Huang
The University of Auckland, New Zealand
k.huang@auckland.ac.nz

Abstract

This study compares the Taiwanese Southern Min (TSM) tones produced by L1 Mandarin speakers and L1 TSM speakers in Taiwan. TSM has seven lexical tones with complex tone sandhi rules while Mandarin only has four lexical tones with a Tone 3 Sandhi. The acoustic study shows that the L1 Mandarin group produced all seven tones with pitch registers and contours similar to the L1 TSM group, including the uncategorized mid-level tone (Tone 7) and two short tones (Tone 4 & 8). In addition, the L1 Mandarin speakers produce accurate sandhi tones at non-XP-final positions. They had no difficulty applying the tone sandhi rules, suggesting that they might have stored the sandhi forms lexically. The further comparison to L1 Mandarin group’s Mandarin tones suggests that the rising and the falling tones were assimilated to their native ones. However, they struggled to reduce the duration of the two TSM short tones to the same degree as the L1 TSM speakers, which suggests that durational differences were more difficult than the pitch for them to acquire.

Index Terms: lexical tones, Southern Min, Mandarin, second language speech, Tone Sandhi, Tone acquisition

1. Introduction

Previous studies have shown that L1 and L2 tonal acoustic characteristics and phonological systems both play an important role in perceiving non-native tones [1-9]. For instance, Cantonese listeners confuse Mandarin high-level tone with high falling tone because the two are allotones in their L1[4, 8]. Furthermore, while naïve listeners might assimilate the L2 tones into their native tone categories based on phonetic similarities [9], the perceptual assimilation occurs at both the phonological and phonetic levels for experienced L2 learners [10]. However, fewer studies have been done on the production of non-native lexical tones by L2 lexical tone users. This study aims to investigate how L1 Mandarin speakers produce Taiwanese Southern Min (TSM) lexical tones.

On the other hand, Mandarin only has four tones that all end with a vowel or a nasal. There are some similarities between TSM and Taiwan Mandarin in high, falling, and low tones. However, it remains unclear whether L1 Mandarin speakers can distinguish the TSM rising from the mid tone. If so, it is unclear which tone will be assimilated to the Mandarin rising tone as studies have shown that the rising tone in Taiwan Mandarin is flatter [11-13]. In addition, there are two “uncategorized” short tones in TSM, and it is worth investigating how L1 Mandarin speakers produce these tones.

Moreover, while Mandarin only has a Tone 3 Sandhi (T3→T2/T3), TSM has a complicated tone sandhi “circle” where the five smooth tones change to another at non-XP-final positions (T3→T2/T1/T7→T3; T5→T3 or T7). For the two short tones, the tone sandhi rules are conditioned by the final obstruent: T4→T1/(-h) and T8→T8/(-p, t, k); T8→T7/(-h) and T8→T4/(-p, t, k). It remains to be seen how L1 Mandarin speakers produce these tone sandhi forms.

This study compares the TSM tone production by L1 TSM and L1 Mandarin speakers. In addition, L1 Mandarin speakers' TSM production is further compared to its Mandarin tone production to examine whether assimilation occurs.

2. Methods

2.1. Materials

The lexical tones of both TSM (7) and Mandarin (4) were elicited in different tonal contexts. 92 TSM and 32 Mandarin disyllabic words were chosen to elicit 49 (7*7) TSM and 16 (4*4) Mandarin bi-tonal combinations, respectively. All the words belong to the colloquial stratum and are frequently used. Both Mandarin and TSM words were written in Chinese characters. The subjects were asked to put Mandarin and TSM words in the frame sentences qǐng shuō__ __bā __ __say three times.’ and guá kóng __ __kóng sann __ __say three times.’ respectively.

2.2. Subjects

20 Mandarin and TSM speakers in Taiwan were recruited. Ten Mandarin-dominant (MD) speakers were Mandarin native speakers who grew up in a Mandarin-speaking family, and their parents did not speak any TSM. These L1 Mandarin speakers have some knowledge of the tested TSM vocabulary as the words chosen were common nouns that are used in daily life in Taiwan, such as food names. The other ten were TSM native speakers (TSM-dominant; TD) who grew up in a family where TSM was the only home language spoken.

2.3. Procedures

The subjects first completed a short questionnaire on their linguistic backgrounds, then read the testing sentences recorded with a Roland R-05 digital recorder with a sample rate of 44100

---

### Table 1: Lexical tones in TSM and Mandarin.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>T1/55/</td>
<td>T1/44/</td>
<td>T1/55/</td>
<td>T1/44/</td>
<td>T1/44/</td>
</tr>
<tr>
<td>Low</td>
<td>T3/21/</td>
<td>T3/21/</td>
<td>T3/21/</td>
<td>T3/31/</td>
<td>T3/21/</td>
</tr>
<tr>
<td>Short</td>
<td>T4/21/</td>
<td>T4/3/</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Rising</td>
<td>T5/24/</td>
<td>T5/12/</td>
<td>T2 /35/</td>
<td>T2 /32/</td>
<td>T2 /33/</td>
</tr>
<tr>
<td>Mid</td>
<td>T7/33/</td>
<td>T7/22/</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Short</td>
<td>T8/5/</td>
<td>T8/5/</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

As shown in Table 1, TSM has seven tones: five “smooth tones” Tone 1, 2, 3, 5, 7 that end with a vowel or a nasal, and two “short tones” Tone 4 and 8 that end with a final obstruent.
Hz in a quiet room. The subjects were asked to repeat each sentence twice at the speed that they were comfortable with.

2.4. Data analysis

The rimes of the target disyllabic words were hand labelled. Durations were extracted, and f0 measurements in semitone were extracted at the midpoint of all the 1/10th intervals with Praat. The pitch contours shown in the figures present the normalized rimes (x-axis) in semitone (y-axis). The data were further analyzed with linear mixed models in SPSS Statistics 25. The fixed effects included position (10 pitch points representing the 5%, 15%, 25%,…, 95% point; last 6 points for tones under stronger carry-over effects) as a covariate, and the position interaction with group, language and/or tones which captured the pitch contour slope differences. The random effects included subjects as random intercepts. By-subject random slopes on tested rime, position and their interactions were also included.

3. Results

3.1. Pitch targets of the lexical tones

3.1.1. High tone and falling tone

The high tone produced by the MD and TD group shared similar contours, as shown in Figure 1.

![Figure 1: TSM high tone by the MD and TD groups.](image1)

Figure 1: TSM high tone by the MD and TD groups.

![Figure 2: TSM falling tone by the MD and TD groups.](image2)

Figure 2: TSM falling tone by the MD and TD groups.

Figure 2 shows the falling tone produced by the two groups. MD’s falling tone seems to have a higher starting pitch and a steeper pitch fall than TD’s. The linear mixed method including tone (high, falling), group (MD, TD), and position as fixed effects was performed. Estimates of fixed effects showed that the starting pitch of the falling tone by TD is 2.21 semitone higher than the high tone by TD (baseline); the differences of the starting pitch between MD and TD were not statistically significant (falling: \( p=0.0537 \); high: \( p=0.940 \)). However, the interaction with position revealed pitch contour differences: both TD and MD’s high tone showed a flat pitch contour and no significant differences \( (p=0.086) \); As for falling tone, TD’s falling tone reduced 0.47 semitone per unit \( (p<0.001) \), and MD’s falling tone further reduced 0.23 semitone per unit, showing a much steeper pitch fall \( (p<0.001) \). The statistical analysis shows that MD’s falling tone had a steeper pitch fall compared to TD.

![Figure 3: TSM and Mandarin falling tone by the MD group.](image3)

Figure 3 shows that the steeper pitch fall by the MD group was similar to their Mandarin high falling tone. The linear mixed methods factoring language, tone, and position revealed that the influence of the language (TSM or Mandarin) was not statistically significant on the starting point \( F(1, 91.195)=0.049, p=0.826 \) or the pitch slope \( (\text{language*position}) \). \( F(1, 2103.323)=2067.212, p=0.468 \). The MD group likely assimilated the TSM falling tone to the L1 Mandarin falling tone.

3.1.2. Low tone

MD and TD shared similar pitch contours for the TSM low tone \( (T3) \), as shown in Figure 4. The linear mixed methods on the last six points of the pitch \( (45\%-95\%) \) factoring group and position revealed insignificant fixed effects from group \( F(1, 16.986)=0.493, p=0.468 \) or group*position \( F(1, 18.286)=0.221, p=0.644 \) were also insignificant.

![Figure 4: TSM low tone by the MD and TD groups.](image4)

Figure 4 shows that the MD group was similar to their Mandarin low falling tone. The linear mixed methods factoring language, tone, and position revealed that the influence of the language \( (\text{TSM or Mandarin}) \) was not statistically significant on the starting point \( F(1, 36.652)=0.001, p=0.974 \) or the pitch slope \( (\text{language*position}) \). \( F(1, 27.488)=0.493, p=0.488 \) were also insignificant.

3.1.3. Rising tone and mid tone

MD and TD’s production of the TSM rising tone and the mid tone are shown in Figures 5 and 6 respectively. The linear mixed model factoring fix effects of tone \( (\text{mid or rising}) \), group \( (\text{TD or MD}) \), position \( (\text{last six points} \ 45\%-95\%) \) and interactions...
revealed significant effects on tone \( F(1, 120.207)=146.485, p<.001 \) and tone-position interaction \( F(2, 150.791)=49.231, p<.001 \), showing that two tones had a different starting pitch point (intercept) and pitch contour (slope). The group-tone interaction was not significant \( F(2, 30.246)=0.244, p=0.785 \), showing that the two groups shared a similar pitch at the 45% point. However, group*tone*position was significant \( F(2, 150.791)=3.622, p=0.03 \). The estimates of fixed effects showed that MD’s rising tone increased 0.21 semitone per unit, but MD’s rising tone further increased 0.16 semitone per unit \( (p=0.011) \). On the other hand, the slope of the mid tone between MD and TD were not different \( (p=0.471) \). To sum up, MD was able to produce distinct mid and rising tone, but the TSM rising tone produced by MD rose steeper than TD’s production.

Figure 5: TSM Rising tone by the MD and TD groups.

I further compared MD’s TSM rising and mid tone with their Mandarin rising tone. The linear mixed model of the three tones revealed significant effects of tone \( F(2, 94.022)=25.275, p<.001 \), position \( F(1, 101.487)=92.062, p<.001 \), and tone–position interaction \( F(2, 101.922)=18.268, p<.001 \). The estimates of the fixed effects showed that at 45% of the rime (intercept), the difference between MD’s Mandarin rising tone and TSM rising tone was not significant \( (p=0.072) \), but the Mandarin rising tone was 1.21 semitone lower than that of the TSM mid tone \( (p<.001) \). The TSM and Mandarin rising tones also shared a rising slope \( (p=0.75) \). However, the TSM mid tone reduced 0.29 semitone from the baseline \( (p<.001) \), showing a flat pitch contour \( (0.34-0.29=0.05 \text{ per unit}) \). The result suggests that MD assimilated the TSM rising tone to their Mandarin rising tone, and they produced the TSM mid tone with a distinct mid-flat pitch contour.

3.1.4. Short tones (Tone 4 & Tone 8)
As shown in Figures 7 and 8, the two short tones produced by the two groups were similar. The linear mixed methods factoring group (TD and MD), tone (Tone 4 and Tone 8) and position confirmed that there were no significant fixed effects from group \( F(1, 18.089)=0.050, p=0.825 \), group*tone \( F(1, 42.049)=1.849, p=0.181 \), group*position \( F(1, 32.795)=0.417, p=0.523 \), or group*tone*position \( F(1, 32.793)=2.402, p=0.131 \). Estimates of fixed effects showed that Tone 8’s starting pitch was 21.24 semitones, while Tone 4 is 1.97 semitones lower. Both tones had a falling pitch contour with pitch reducing -0.28 semitone per unit. There were no significant differences between the pitch contour slopes of the two tones \( (p=0.532) \).

Figure 7: TSM Tone 4 by the MD and TD groups.

Figure 8: TSM Tone 8 by the MD and TD groups.

3.2. Tones at non-XP-final positions
Figure 9 shows the comparison of the original Tone 1, 2, 3, 7 at non-XP-final positions. Tone 5 was not included due to variations of its sandhi forms.

![Figure 9: TSM Tone 1, 2, 3, 7 at non-XP-final positions by the MD and TD groups.](image-url)
unit) were steeper than those by TD (-0.45 semitone per unit) (p<0.001), and sandhi Tone 1 (low) by MD were flatter (-0.10 semitone per unit) than those by TD (-0.21 semitone per unit) (p=0.003).

![Figure 10: TSM Tone 1, 2, 3, 7 at non-XP-final positions vs Mandarin tones by the MD group.](image)

A similar within-MD comparison was made between their four sandhi tones in TSM and the four Mandarin lexical tones at non-XP-final positions (Figure 10). The sandhi Tone 1 (mid) was compared to the Mandarin rising tone. Estimates of fixed effects showed that the starting pitch intercepts were different for high, low, and rising tones between TSM and Mandarin (p=0.001, p<0.001, p<0.001 respectively). However, only MD’s Mandarin rising tone (+0.41 semitone per unit) and sandhi Tone 1 (mid) (+0.01 semitone per unit) showed a different contour (p<0.001). This suggests that MD speakers did not assimilate the sandhi Tone 1 (mid) to the Mandarin rising tone. They were aware of the difference between mid and rising tone at non-XP-final positions, even though a rising pitch contour was not found at the position in their L1.

![Figure 11: TSM short tones at non-XP-final positions by the MD and TD groups.](image)

Figure 11 compares the sandhi Tone 4 and Tone 8 produced by both groups. The analysis included tone, group*tone, tone*position, and group*tone*position as fixed effects. Estimates of fixed effects revealed that when the final was -p, -t, or -k, there were no significant differences between TD and MD for the tone intercepts of both sandhi Tone 4 (p=0.851) and sandhi Tone 8 (p=0.361). There were also no significant differences on the tone slopes between TD and MD for both sandhi Tone 4 (p=0.428) and Tone 8 (p=0.568). When the final was -h, there was no significant influence of the group on tone—both TD and MD shared similar starting pitch registers on both tones. For the sandhi Tone 8 with -h, there was also no significant difference on the pitch fall (MD: -0.39; TD: -0.45 per unit; p=0.263). However, the sandhi Tone 4 with -h had a steeper pitch fall for MD (-0.63 semitone per unit), compared to TD (-0.44 per unit). The difference was statistically significant (p=0.002).

### 3.3. Durations

The two TSM short tones (Tone 4 and Tone 8) are shorter than the five smooth tones. The result shows that the MD speakers produced shorter durations for the short tones. The rise durations of the entering tones that end with final /p, t, k/ were reduced to 57.2% of the average duration of the smooth tones. However, the reduction was not as much as the TD group (47.9%). The differences between the two were statistically significant (t(516)=5.74, p<0.001).

As for short tones with -h, the result shows that neither group reduced the durations at non-XP-final positions (92.1% for MD and 93.6% for TD), as the sandhi form is a smooth tone. However, when the tone sandhi does not apply, MD speakers did not reduce the duration (M=0.872), while TD speakers slightly reduced the durations (M=0.815). The differences between the two groups were statistically significant (t(534)=3.03, p=0.003)

### 4. Discussion

The results suggest that overall, MD and TD produced similar pitch contours of TSM high and low tones. As for TSM falling tone, the two groups shared a similar pitch register, but the MD group had a steeper pitch fall, which was likely a result of assimilation to their L1 Mandarin falling tone. The results of these three tones also confirm the previous observations [11, 15] that the starting pitch of the falling tone is higher than the high tone in both TSM and Taiwan Mandarin.

The results of TSM rising and mid tones suggested that the L1 Mandarin speakers assimilated the TSM rising tone to the Mandarin rising tone, while they successfully acquired the mid tone even at non-XP-final positions. In PAM’S term [3], TSM mid and rising were likely perceived as ‘Uncategorized-Categorized’ respectively, with no perceived overlap. The results also further suggest that for the MD speakers, the rising contour remains to be the most salient feature of Mandarin Tone 2, which contradicts previous findings that suggested Mandarin Tone 2 is dipping or mid [11-13]. Otherwise, we might have seen them treat TSM mid tone as the ‘categorized’.

As for the TSM short tones, the L1 Mandarin speakers successfully produced the pitch register and contour similar to the native speakers. However, they failed to reduce the duration to the degree of the native speakers, suggesting that the duration of the prosody poses more challenges for them.

Lastly, the MD group performed well in producing sandhi tones. This is likely due to the experiment design because all the elicited vocabulary were common words. The L1 Mandarin speakers had likely lexicalized the sandhi forms and treated them as a different tone. Future studies need to include sandhi forms in novel situations to see if the speakers truly acquire the tone sandhi rules.

### 5. Conclusions

The results show that L1 Mandarin speakers in Taiwan successfully acquired the uncategorized mid-level tone and short tones in TSM, and they were able to produce similar pitch contours of the sandhi tones as well. However, acquiring durational differences in lexical tones was more challenging for them.
6. References


