The effect of Min proficiency on production and perception of tones in Taiwan Mandarin

Yi-Hsuan Huang, E-chin Wu & Janice Fon*

Graduate Institute of Linguistics, National Taiwan University, Taiwan

d97142003@ntu.edu.tw, jfon@ntu.edu.tw

Abstract

This paper aimed to investigate how Min proficiency affects tonal production and perception in Taiwan Mandarin. Experiment 1 recruited two groups of Mandarin-Min bilinguals, with Min proficiency varied. Results showed that high Min proficiency speakers produced a lower tonal register and a narrower tonal range in Mandarin compared to low Min proficiency speakers. The tonal variations also interacted with tonal target types. Experiment 2 further investigated whether different levels of pitch heights led to any difference in listeners' interpretation of talkers' language backgrounds. Results showed that the lower the tonal register was, the more likely the listeners regarded the talker as a "Min-speaking" person, mirroring the results on tonal production.

Index Terms: tonal variation, Taiwan Mandarin, Southern Min, indexical properties, bilingualism

1. Introduction

Taiwan is a multilingual society where the major speech community is composed of Mandarin-Southern Min (Min hereafter) bilinguals [1]. Putonghua Mandarin has four lexical tones, high-level, high-rising, mid-dipping, and high-falling, represented by tonal letters of 55, 35, 214, and 51, respectively in [2]. Taiwan Mandarin is described as also having four tones, which include high-level, mid-dipping, low-dipping, and high-falling [5], characterized as 44, 323, 312, and 42. Min, however, has five long tones, including high level, low-rising, low-dipping, high-falling, and mid-level tones that were characterized by tonal letters of 44, 24, 31, 52, and 33 in [3].

Although Mandarin is the official language in Taiwan, more than 70 percent of the population speak Min to various degrees [4]. Min is thus considered to play an important role in shaping tonal realizations in Taiwan Mandarin [5-7]. Compared to Mandarin, Min is a tone language that prefers a lower pitch register [8]. It is proposed that the tonal lowering effects found in the registers of Taiwan Mandarin can be accounted for by the influence of Min [5, 7]. Since empirical studies on this issue have been scarce, Experiment 1 aims to investigate how Min proficiency interacts with tonal realizations in Taiwan Mandarin by using a controlled experimental paradigm. Our prediction is that the better one's Min is, the more his or her tonal register in Mandarin is pulled down by Min.

Furthermore, if there is any pitch variation associated with speakers' Min proficiency in tonal production, it would also be interesting to see whether such a tonal variation is still discernible on the perceptual side. If Mandarin tonal variation is influenced by Min proficiency, it is plausible that tonal register in Mandarin might provide some indexical cues that indicate the talker's language background [9, 10] (i.e., Min proficiency in the current case). Experiment 2 aims to investigate this issue. We predict that the lower the tonal register is, the more likely listeners would associate it with speakers' fluency in Min.

2. Methods

2.1 Experiment 1

2.1.1 Participants

Forty Mandarin-Min bilinguals, balanced in gender, participated in this study. They were born and raised in the Taipei metropolis and spoke Mandarin as their dominant language. All of them, including their parents, were ethnically Min. They were further divided into two groups, the high proficiency and the low proficiency group based on phone interviews by the experimenter prior to the experiment. In the interview, participants were asked to introduce themselves in Min briefly and to answer experimenters' questions. The experimenter then provided rating scores based on participants' fluency, vocabulary use, and pronunciation in Min.

2.1.2 Stimuli

Stimuli were 10 monosyllabic quadruplets that covered the four lexical tones in Mandarin. Syllables within each set only differed from one another in tone. All the syllables were of either CV or CVN structures.

2.1.3 Equipment

Stimuli were recorded by a SONY PCM-M1 Digital Audio Recorder and a SHURE SM10A head-mounted microphone.

2.1.4 Procedure

The experiment was conducted in a sound-proof booth. Stimuli were written in traditional Chinese characters and were presented to participants using Microsoft Office PowerPoint 2003 on a laptop. Each slide contained one stimulus. Participants were asked to read aloud the character presented on the screen as natural as possible at a normal speed. They were asked to re-read the tokens if incorrect or unnatural pronunciations were produced.

2.1.5 Measurements

The sound files were sliced and the voiced portions of the syllable were hand-labeled using Praat [11]. A Praat script was written for pitch extraction, which was then checked and corrected manually to avoid pitch doubling or halving. A second script was written to extract ten pitch points of equal time intervals between the onset and the offset of the voiced portion. A third script was written to extract the reference points for the stimuli of each tone. For Tone 1, the F_0 maximum of each syllable was extracted. For dipping tones such as Tone 2 and full Tone 3, the initial F_0 maximum, the

medial F_0 minimum, and the final F_0 maximum were extracted as reference points for statistical analyses. For tonal contours that were realized as falling shapes, such as Tone 4 and partial Tone 3, the initial F_0 maximum and the final F_0 minimum were measured.

2.2 Experiment 2

The goal of Experiment 2 is to investigate whether tonal variations in pitch height can help listeners identify talker's linguistic backgrounds.

2.2.1 Participants

Thirty Mandarin-Min bilinguals, 14 males and 16 females, participated in this experiment. All were ethnically Min, including their parents. They were all proficient Min speakers. According to their self-rating scores using a 7-point Likert scale, they had a mean proficiency of 5.8 ($SD = 0.82$).

2.2.2 Stimuli

Thirty-six high-frequency disyllabic words covering all tonal combinations in Mandarin were used in this experiment. Stimuli with Tone 3 occurring in the second syllable were eliminated in the design since utterance-final Tone 3s tended to be accompanied by creakiness, which caused the results of acoustic manipulation to sound unnatural. For each disyllabic word, three levels of pitch height, low, mid, and high, were produced by a female Mandarin-Min bilingual. Later, pitch register was adjusted by Praat so that initial pitch jumps from low to mid, and from mid to high were of equal steps, as was shown in Table 1.

Table 1. *Target onset pitch (in Hz) for each tone.*

Levels	Tone 1	Tone 2	Tone 3	Tone 4
Low	220	190	180	230
Mid	250	220	210	260
Low	280	250	240	290

2.2.3 Equipment

The equipment was the same as that used in Experiment 1.

2.2.4 Procedure

The audio stimuli were played to participants by Microsoft Office PowerPoint 2003, with the speed of presentation controlled by the experimenter. For each trial, participants first listened to the target stimulus and were asked to decide whether they believe the talker speaks Min or not using a two-alternative forced choice paper-and-pencil test. Each level of a stimulus (with high, mid, and low pitches) only occurred once in the experiment.

2.2.5 Measurements

Participants' answers to each trial were coded as "0" (yes) and "1" (no). A chi-square test was performed to see if the responses were affected by different levels of tone heights.

3. Results

3.1 Experiment 1

Average F_0 trajectories across tones produced by high Min proficiency (HMP) and low Min proficiency (LMP) speakers are shown in Figure 1. For Tone 1, a planned independent t-test was conducted to see whether Min proficiency had any effects on the F_0 maximum of Tone 1 within each gender. Results showed that HMP male speakers significantly produced lower F_0 maximum for Tone 1 compared to LMP male speakers [$t(188) = -2.84$, $p < .01$]. The result was also significant for female speakers [$t(198) = -3.50$, $p < .01$].

For Tone 2, the general trend was that the HMP speakers had lower F_0 for the three reference points of Tone 2, which were the initial F_0 maximum, the medial F_0 minimum, and the final F_0 maximum. A Proficiency (2) × Point (3) mixed ANOVA was carried out for both gender groups to confirm this observation. For males, both the main effects of Proficiency and Point were found [Group: $F(1, 195) = 42.19$, $p < .0001$, $\hat{\eta}^2 = 0.18$; Point: $F(1, 195) = 132.08$, $p < .0001$, $\hat{\eta}^2 = 0.40$]. The two-way interaction was not significant. For Proficiency, post hoc independent t test showed that speakers in the HMP group showed lower F_0 compared to those in the LMP group [$t(589) = -8.18$, $p < .0001$] across all three reference points. For Point, post hoc Tukey's b test showed that the final F_0 maximum was the highest while the medial F_0 minimum was the lowest, with the initial F_0 maximum in between. For females, results showed that both the main effects of Proficiency and Point were significant [Proficiency: $F(1, 196) = 23.87$, $p < .0001$, $\hat{\eta}^2 = 0.11$; Point: $F(2, 392) = 849.27$, $p < .0001$, $\hat{\eta}^2 = 0.81$]. The two-way interaction was also significant [Proficiency × Point: $F(2, 392) = 8.46$, $p < .0001$, $\hat{\eta}^2 = 0.04$]. Post hoc independent t test showed that female HMP subjects significantly had lower F_0 than female LMP subjects across all three reference points [I: $p < .0001$, M: $p < .01$, F: $p < .0001$], with the group difference being less distinct in the medial F_0 minimum.

Since Tone 3 in Taiwan Mandarin can be realized as either dipping or low-falling, these two realizations were categorized and analyzed separately. For dipping Tone 3, a Proficiency (2) × Point (3) mixed ANOVA was conducted for both gender groups to examine the effect of Min proficiency on the F_0 contour for Tone 3. For males, both of main effects were found [Proficiency: $F(1, 162) = 272.85$, $p < .0001$, $\hat{\eta}^2 = 0.22$; Point: $F(2, 162) = 272.85$, $p < .0001$, $\hat{\eta}^2 = 0.77$]. The two-way interaction was also significant [Proficiency × Point: $F(2, 162) = 3.99$, $p < .05$, $\hat{\eta}^2 = 0.05$]. Post hoc independent t tests showed that the LMP speakers had higher F_0 than the HMP speakers across all three positions [initial & medial: $p < .0001$, final: $p < .01$], with the group difference being greater in the initial and the medial positions than in the final position. For females, only the main effect of Position was found [Position: $F(2, 186) = 270.18$, $p < .0001$, $\hat{\eta}^2 = 0.74$]. The two-way interaction was significant [Proficiency × Position: $F(2, 186) = 17.89$, $p < .0001$, $\hat{\eta}^2 = 0.16$]. Post hoc independent t test showed that the HMP had lower F_0 than the LMP ($p < .01$) for the initial F_0 maximum. However, for the other two points, the trend was the opposite, with the LMP subjects showing lower F_0 than the HMP subjects [medial: $p < .01$, final: $p < .05$].

For low-falling Tone 3, a Proficiency (2) × Point (2) mixed ANOVA was conducted to see the effect of Min proficiency on the initial F_0 maximum and the final F_0 minimum. For males, results showed that the main effect of Point was found [Point: $F(1, 115) = 868.3$, $p < .0001$, $\hat{\eta}^2 = 0.88$]. The two-way interaction was also significant [Proficiency × Point: $F(1, 115) = 24.18$, $p < .0001$, $\hat{\eta}^2 = 0.17$]. Post hoc independent t test showed that the HMP group

significantly produced lower F_0 than the LMP subjects for the initial F_0 maximum ($p < .0001$). No proficiency effect was found for the final F_0 minimum. For females, results showed that the main effect of Point was found [Position: $F (1, 103) = 1250.35$, $p < .0001$, $\hat{\eta}^2 = 0.92$]. The two-way interaction was also significant [Proficiency × Point: $F (1, 103) = 10.97$, $p < .01$, $\hat{\eta}^2 = 0.10$]. Post hoc independent t test showed that the HMP subjects had significantly lower F_0 than the LMP subjects for the initial F_0 maximum ($p < .01$). No proficiency effect was found for the final F_0 minimum.
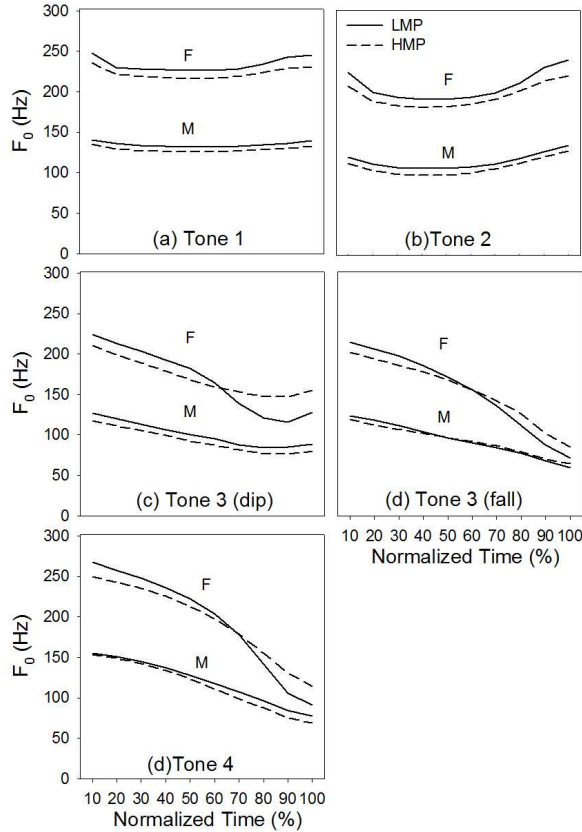

Figure 1: *Time normalized pitch contours in high Min proficiency (HMP) and low Min proficiency (LMP) groups for males (M) and females (F).*

For Tone 4, a Proficiency (2) × Point (2) mixed ANOVA was conducted for both genders. For males, results showed that both main effects of Proficiency and Point were found [Proficiency: $F (1, 198) = 22.9$, $p < .0001$, $\hat{\eta}^2 = 0.10$; Point: $F (1, 198) = 1678$, $p < .0001$, $\hat{\eta}^2 = 0.89$]. The two-way interaction was not found. For Proficiency, post hoc independent t test showed that the HMP group had lower F_0 than the LMP group, regardless of reference points [$t (198) = -4.25$, $p < .001$]. For Point, post hoc paired t test showed that the initial F_0 maximum was significantly higher than the final F_0 minimum [$t (199) = 41.04$, $p < .0001$]. For females, a Proficiency (2) × Position (2) mixed ANOVA was carried out to see if there was any group difference at each reference point. Results showed that the main effect of Position was significant [Position: $F (1, 198) = 1467.15$, $p < .0001$, $\hat{\eta}^2 = 0.88$]. The two-way interaction was also significant [Group × Position: $F (1, 198) = 25.71$, $p < .0001$, $\hat{\eta}^2 = 0.12$]. Post hoc independent t

test showed that the LMP subjects had higher initial F_0 maximum than the HMP subjects [$t (199) = 41.04$, $p < .0001$]. As for the final F_0 minimum, however, the trend was in the opposite direction, with the LMP subjects having lower F_0 than the HMP subject [$t (199) = -2.98$, $p < .05$].

3.2 Experiment 2

Figure 2 illustrates the results for Experiment 2, showing that the higher the target pitch level is, the less likely the talker was regarded as a "Min-speaking person" by listeners. A chi-square test of goodness of fit was performed to see whether listeners' responses (yes vs. no) were related to the target pitch levels or not. Results showed that Level had significant influences on listeners' response categories [$\chi^2 (2, N = 3240) = 47.68$, $p < .0001$]. Specifically, when listening to stimuli with a low pitch height, subjects were more likely to associate Min fluency with the talker. Importantly, the percentage of "yes" responses gradually decreased when they encountered stimuli with higher levels of target pitch heights.
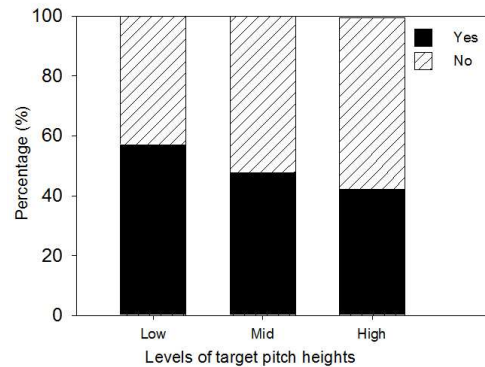

Figure 2: *The percentage of yes-no responses for the question "Do you think the talker speaks Min or not?" across stimuli of three levels of target pitch heights.*

Although most listeners in this study considered themselves proficient Min speakers, with most of the self-rating scores above 5, some only gave Score 3 or Score 4 to their own Min proficiency. To eliminate potential confounding factors resulting from different degrees of Min proficiency, we further divided the subjects into two groups based on their average self-rating scores on Min proficiency in speaking and listening. Subjects whose average Min proficiency scores exceeding or equaling 5 were categorized as the "high Min proficiency group" (HMP) while those who had average ratings between 4 and 5 were classified as the "mid Min proficiency group" (MMP).

A binary logistic regression analysis was then conducted to see if Level (3) and listeners' Min Proficiency (2) served as predictors in listeners' judgments of whether the talker could speak Min or not. A test of the full model against a constant-only model was statistically significant, indicating that the predictors as a set could significantly distinguish between "Min-speaking" and "not Min-speaking" responses [$\chi^2 (2) = 48.76$, $p < .0001$]. The Wald criterion demonstrated that both Level and listeners' Proficiency made significant contribution to the prediction (Level: $p < .0001$, Proficiency: $p < .05$). For Level, the Exp(B) value indicated that the odds ratio for stimuli with the pitch level "Low" was 1.803 compared to the pitch level "High". Furthermore, the odds ratio for the pitch level "Mid" was 1.229 compared to the pitch level "High".

For Proficiency, the Exp(B) value indicated that the odds ratio for the HMP group was 0.764 compared to the MMP group, showing that listeners who were highly proficient in Min tended to consider the talker as "not Min-speaking".

4.   Discussion & Conclusion

Experiment 1 compared the realization of four lexical tones in Taiwan Mandarin between high Min proficiency and low Min proficiency speakers from the production side. The aim was to see whether Min proficiency imposed any effects on tonal realization in Mandarin. Results showed that high Min proficiency speaker demonstrated a lower pitch register, a narrower pitch range, and a less drastic contour change compared to low Min proficiency speakers. Since these two groups of speakers came from the same dialectal area but only differed in their degrees of Min proficiency, it is plausible that such tonal variations are related to their fluency differences in Min, confirming our hypothesis that speakers' tonal registers or tonal ranges in Mandarin might be pulled down or compressed by Min if they can speak Min fluently. Experiment 1 attested Fon et al.'s [5] proposal, which indicated that the tonal realization of Taiwan Mandarin might be influenced by Min and further proved that such influences are consistently correlated with Min proficiency.

In addition, another interesting phenomenon worth noticing was that the direction for tonal variations seemed to interact with the types of tonal targets. For tones with high tonal targets, speakers with high Min proficiency produced lower F_0 than speakers with low Min proficiency. However, for tones with low tonal targets, such as Tone 3 or the final point of Tone 4, the direction of tonal variation was reversed. Instead, it was the proficient Min speakers that demonstrated higher F_0 for low tonal targets compared to non-proficient Min speakers. It turned out that proficient Min speakers distinguished their ways of tonal realization in Mandarin from non-proficient Min speakers through the lowering of the high tonal targets and the raising of the low tonal targets, leading to the impression of a narrower tonal range and a flatter tonal contour. This study showed that tonal realization in Mandarin for proficient Min speakers was not an overall lowering on their pitch registers but was dependent on target types. Tonal variation was robust for high targets while subtle for low targets. Such an interaction between tonal realizations and target types indicated that these variations were not realized monotonically, consistent with [12]. Instead, it is crucial to take the types of tonal targets into consideration.

Now that the tonal effects stemming from Min was found to be implemented by tonal heights on the production side, Experiment 2 further tested whether or how the pitch height differences in Mandarin led to different projections of the talker's language background, namely, Min proficiency. Results showed that levels of pitch heights successfully served as a significant predictor for listeners' responses, indicating that pitch height was influential for listeners' interpretations of talkers' Min proficiency. There seemed to be a positive correlation between the degree of pitch lowering and the probability of "Min-speaking" responses. Particularly, the lower the pitch register was, the more likely listeners would judge it as a token produced by a Min-speaking person. Our findings proved that tonal variations on the production side were not merely due to gradiency in acoustic signals, it was also detectable by listeners, serving as one of the indexical properties of talkers and providing evidence for the interplay of tonal heights and Min proficiency on the perception side in Taiwan Mandarin.

Though the influence of pitch levels on judgment responses was in line with our prediction, listeners' Min proficiency was found to be one of the predictors as well. It is plausible for listeners to perceive different dialects based on their language background [10]. However, it was surprising that when presented with Mandarin stimuli, listeners with high Min proficiency tended to regard the talker as a "non-Min-speaking" person compared to listeners with a mid-level of Min proficiency. One explanation could be that proficient Min listeners might have stringent criteria for a person to be recognized as "Min-speaking". They might be paying extra attention to other acoustic details which might also be reflective of one's language background, such as segmental information. For less proficient Min listeners, they probably could not fine-tune their perception system as accurately as the more proficient listeners. It would thus be more difficult for them to detect any other possible cues that shape the characteristics of a different dialect. Another reason may be due to the language backgrounds of our female speaker. She was a Mandarin-Min bilingual and could speak Min fluently, but one of her parents was a Mandarin monolingual, which may further complicate bilinguals' judgments for the indexical properties encoded in her voice.

In sum, proficient Min speakers had lower tonal registers, narrower tonal ranges in Taiwan Mandarin than non-proficient Min speakers. The tonal variations also interacted with types of tonal targets. On the perception side, listeners tended to associate lower tonal registers with Min-speaking abilities.

5. References

[1]   D. C. S. Li and S. Lee, "Bilingualism in East Asia," in *The handbook of bilingualism* T. K. Bhatia and W. C. Ritchie., Eds., ed Malden, Mass: Blackwell Press, 2006, pp. 742-779.

[2]   Y.-R. Chao, *A Grammar of Spoken Chinese*. Berkeley & Los Angeles: University of California, 1968.

[3]   R. L. Cheng, "Tone Sandhi in Taiwanese," *Linguistics,* vol. 41, pp. 19-42, 1968.

[4]   S. Huang, *Yuyan, shehui yu zuqun yushi-- Taiwan yuyan shehuixue de yanjiu [Language, society, and ethnic identity--Studies in language socialogy in Taiwan]*. Taipei: Crane, 1993.

[5]   J. Fon and W.-Y. Chiang, "What does Chao have to say about tones?--a case study of Taiwan Mandarin," *Journal of Chinese Linguistics,* vol. 27, pp. 15-37, 1999.

[6]   J. Fon*, et al.*, "Production and perception of the two dipping tones (Tone 2 and Tone 3) in Taiwan Mandarin," *Journal of Chinese Linguistics,* vol. 32, pp. 249-281, 2004.

[7]   Y.-H. Huang and J. Fon, "Dialectal variations in tonal register and declination pattern of Taiwan Mandarin. ," in *The 4th International Conference on Speech Prosody*, Brazil, 2008, pp. 605-608.

[8]   S. H. Chen, "The effects of tones on speaking frequency and intensity ranges in Mandarin and Min dialects," *Journal of Acoustical Society of America,* vol. 117, pp. 3225-3230, 2005.

[9]   D. Abercrombie, *Elements of general phonetics*. Chicage: Aldine, 1967.

[10] C. G. Clopper, "Effects of talker variability on perceptual learning of dialects," *Language and Speech,* vol. 47, pp. 207-239, 2004.

[11] P. Boersma and D. Weenink, "Praat: doing phonetics by computer," 5.2.23 ed, 2011.

[12] Y. Huang and J. Fon, "Investigating the effect of Min on dialectal variations of Mandarin tonal realization," in *The 17th International Congress of Phonetic Sciences*, Hong Kong, China, 2011, pp. 918-921.