DIALECTAL DIFFERENCES IN THE REALISATION OF TONE 1 IN TAIWAN MANDARIN

Janice Fon[1] & Yu-Ying Chuang[2]

[1]National Taiwan University, Taiwan; [2]Eberhard-Karls University of Tübingen, Germany

jfon@ntu.edu.tw; yychuang2012@gmail.com

ABSTRACT

The central dialect of Taiwan Mandarin is said to have an utterance-final rise. Previous literature indicated that the low-fall Tone 3 and the high-fall Tone 4 in the standard northern dialect are realised respectively as low-dipping and high-level utterance-finally in the central variety. This study thus explored whether such a tendency also extended to the high-level Tone 1. Twenty-five native speakers (13 northern and 12 central) were recruited to read 24 bisyllabic stimuli containing Tone 1. Results showed rising was indeed the predominant realisation for Tone 1, but it was not exclusive to the central dialect, nor was it to the utterance-final position. Reference points for the rising portion also did not show a more drastic rise for the central variety. Rather, central speakers consistently realised their Tone 1 with a steeper initial fall before the rise, likely creating a percept of a more prominent rise than their northern counterparts.

Keywords: Dialectal difference, Tone 1, Level tone, Tonal realisation, Taiwan Mandarin.

1. INTRODUCTION

Mandarin is the official language of Taiwan, a small island country in Southeast Asia. It is a tone language with four lexical tones, a high-level Tone 1, a mid-rising Tone 2, a low-falling Tone 3, and a high-falling Tone 4 [1, 2]. Although the language is vibrant with phonetic variations due to frequent contact with Min, a substrate language and a former *lingua franca* [3], most are in fact nontonal [4, 5]. When it comes to tonal variations, however, many would unanimously point to the central dialect [6].

Two traits are often mentioned by the general public [6]. First, the central dialect favours a lower pitch register and a narrower pitch range than the northern standard dialect, and the effect is especially prominent on high targets [6, 7]. Min proficiency might be an underlying cause for pitch register lowering [7], and pitch range narrowing likely stemmed from a negative transfer from the local Min variety, in which the phonologically high-level tone is realised as phonetically mid-level instead [6].

Secondly, the central dialect tends to have an utterance-final rise [6]. Although this trait is often mentioned in the mass media, systematic research is rare. To the best of our knowledge, there are only two relevant studies. Fu [8] found that junior high schoolers from central Taiwan tend to realise their utterance-final Tone 3 as dipping, instead of the default low fall. Similarly, Wu [9] found that adult central speakers are inclined to realise their utterance-final Tone 4 as high level, rather than the prescribed high fall. Both studies adopted subjective judgement as a means for tonal shape identification, and acoustic analyses are yet to be done.

To evaluate whether the preference for a final rise is specific to only Tone 3 and Tone 4 or common across all four tones, this study extended the inquiry to Tone 1, which would potentially be an ideal candidate since it is realised with lower pitch in the central dialect [6, 7], and any preference for ending with a rise could thus be readily detected.

This study has two aims. First, we would like to use acoustic measures to see if Tone 1 indeed tends to be realised with a rise in the central variety. Previous studies [8, 9] used only subjective judgments, which are efficient for determining phonological categories, but difficult to detail phonetic differences. This study would thus like to complement subjective judgments with acoustic measurements so as to provide a fuller view of the dialect. Secondly, we intend to directly compare the central variety with the northern dialect. Although studies regarding both varieties are existent (cf. northern: [1, 4, 5, 10, 11]; central: [8, 9]), few have included both for direct comparisons, except for Fon [2], which focused on tonal contour, and Huang & Fon [7] and Khoo [6], which focused on tonal range and register. None has examined the utterance-final rise. Therefore, we included both dialects to directly gauge the degree of divergence between the two.

2. METHOD

2.1. Subjects

Twenty-five native Mandarin speakers aged between 18 and 25 were recruited in this study ($\overline{X}$= 20.72, *SD* = 1.97). About half were from northern Taiwan (*N* =

13), and half from central. The two groups were comparable in age [$t(23) = -0.07$, ns]. All subjects were born and raised in their respective areas and had not lived outside the region for more than six months before 18 years old. Each group was about equally divided in gender.

2.2. Stimuli

Three Tone 1 syllables, *ban*[1], *dan*[1], and *gan*[1], were chosen as target stimuli and were paired with another syllable to form disyllabic words. The target stimuli were placed in either the first or the second position, and the pairing syllables were in one of the four tones (e.g., ***dan***[1]*xin*[1] 'to worry' *vs. ming*[2]***dan***[1] 'roster'). In total, there were 3 (syllables) × 2 (positions) × 4 (tonal environments) = 24 stimuli. An additional set of 72 bisyllabic fillers that did not contain any of the target syllables was also included.

2.3. Equipment

Recording was done with a sampling rate of 48 kHz using a SONY PCM-M1 Digital Audio Recorder and a SHURE SM10A head-mounted microphone, and was later downsampled to 22050 kHz using Praat 6.1 [12].

2.4. Procedure

Subjects read the stimuli in a clear and natural fashion in a quiet room. They were randomly assigned to one of the three semi-randomization orders. Words containing the same target syllable were hand-adjusted to avoid juxtaposition.

2.5. Analyses

The voiced portions of the stimuli were hand-labelled using Praat [12], and pitch contours extracted by a Praat script were interpolated and smoothed after hand correction for pitch-doubling and -halving. Tonal shapes of the stimuli were hand-labelled based on independent perceptual judgments on the pitch excursions by three native speaker judges, two of whom were the authors.

To facilitate comparisons across different tokens, two sets of pitch information were extracted for further analyses. The first included reference points from each token (cf. [1, 6, 7, 10]). For the rising contour, it is the initial minimum and final maximum of the rising portion. The second set extracted ten pitch points at equal time intervals from each token, so that analyses could be performed on time-normalised tonal contours.

3. RESULTS

In total, 24 (stimuli) × 25 (subjects) = 600 tokens of stimuli were collected, among which 439 were perceived as rising by at least two judges, accounting for 73% of the data (Table 1). In other words, rising is the most prevalent for both dialects, and all speakers adopted it to various extents. Fig. 1 shows the pitch excursion for the perceived level and rising contours for both genders. The two tonal excursions occupied similar pitch registers, and rising was realised as slightly dipping acoustically, which is common in Taiwan Mandarin [1, 2, 10].

Table 1: Perceived tonal contours of the target stimuli by at least two of the three judges. Numbers before the slashes are counts for the first syllable and those after are counts for the second.

	Level	Rising	Others	Total
Northern	48/12	106/142	2/2	156/156
Central	60/18	71/120	13/6	144/144
Total	108/30	177/262	15/8	600

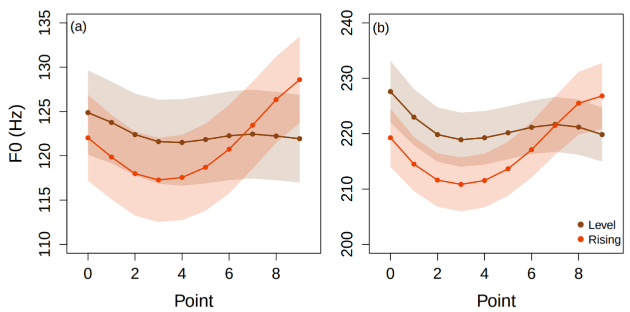

Figure 1: Time-normalised mean pitch excursions of perceived level and rising contours in (a) male and (b) female speakers. Shaded areas represent standard error.

To study how the distribution of rising contours was affected by various factors, a GENDER × DIALECT × POSITION × TONE four-way mixed ANOVA was performed on the rising percentages of each subject. Results showed that all main effects except for DIALECT were significant [GENDER: $F(1, 21) = 4.86$, $p < .05$; POSITION: $F(1, 21) = 28.37$, $p < .0001$; TONE: $F(3, 63) = 5.92$, $p < .01$]. The interaction between POSITION and TONE was also significant [$F(3, 63) = 2.82$, $p < .05$]. Post hoc analyses using Bonferroni's adjustments showed that females had more rising realisations than males (Fig. 2), and final syllables had more rising than prefinals (Fig. 3). With regard to the interaction, there was an effect of TONE for the prefinal position. Tone 1 pairing syllables elicited fewer rising realisations than Tone 2 ($p = .06$) and Tone 3 ($p < .05$), as shown in Fig. 3. No effect involving TONE was found for the final position.

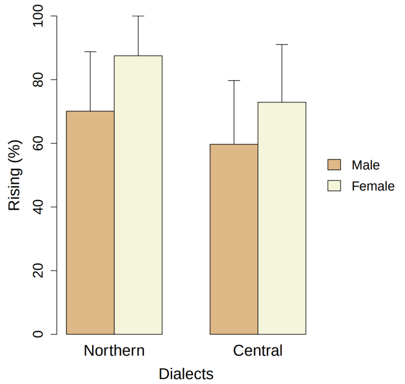

Figure 2: Rising percentages regarding GENDER and DIALECT. Error bars represent standard error.

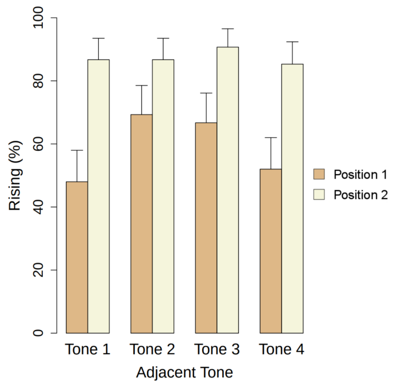

Figure 3: Rising percentages regarding POSITION and TONE. Error bars represent standard error.

Fig. 4 shows the mean pitch of the initial and final points of the rise. Two separate linear mixed effects analyses were performed on males and females using the `lme4` package [13] in R [14], as shown in (1). Reference pitch values were included as the dependent variable, and fixed effects of DIALECT, POSITION, and POINT, along with their interaction terms, were entered into the models. Random effects included by-subject and by-item intercepts, as well as the by-subject random slopes for POINT. Significance was calculated using normal approximation [15].

(1) F0 ~ DIALECT * POSITION * POINT
+ (1+POINT|subject) + (1|item)

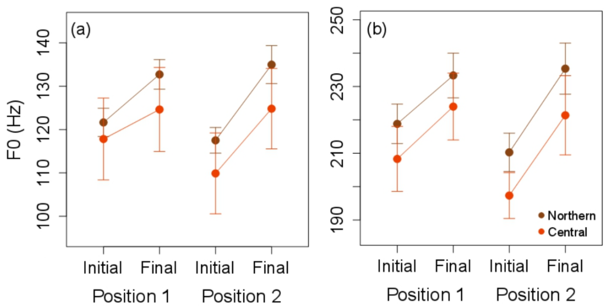

Figure 4: Mean pitch of the initial and final points of the rise across the two dialects for (a) males and (b) females.

Table 2 shows the fixed effects for males and females. For both genders, the main effects of POSITION and POINT were significant. This was expected, since reference points were extracted from the rising portion of a tone, and there was a slight declination effect for the initial reference point. A two-way interaction of POSITION × POINT was also significant. The second syllable consistently showed a larger rise than the first.

Table 2: Fixed effects of the linear mixed models for males and females using the initial point of the first position in the northern dialect as the reference. '**' $p < .01$, '***' $p < .001$.

	Est.	SE	*t* value	
		Male		
Intercept	109.86	6.74	16.29	***
Dialect2	-0.68	9.56	-0.07	
Position2	-9.56	3.30	-2.90	**
Point2	10.86	2.15	5.06	***
Dialect2:Position2	-5.68	4.54	-1.25	
Dialect2:Point2	-4.28	3.19	-1.34	
Position2:Point2	6.67	1.90	3.52	***
D2:Pn2:Pt2	2.36	2.86	0.83	
		Female		
Intercept	203.86	6.04	33.76	***
Dialect2	-8.22	8.92	-0.92	
Position2	-19.00	4.42	-4.30	***
Point2	14.49	3.06	4.73	***
Dialect2:Position2	-4.48	6.21	-0.72	
Dialect2:Point2	-1.08	4.72	-0.23	
Position2:Point2	10.84	2.49	4.36	***
D2:Pn2:Pt2	0.41	3.92	0.11	

Fig. 5 shows the 10-point pitch extractions of the rising realisations from both dialects. To compare the tonal contours, the generalised additive mixed model (GAMM), a powerful statistical method that could mathematically separate tonal undulations into pitch height and pitch excursions [16], was adopted as a means of analyses to accommodate the fact that the contours were not linear. Separate GAMM models were built on males and females using the first position of the northern dialect as the reference to fit the pitch extraction values using the `bam` function of the `mgcv` package [17] in R [14], as shown in (2).

(2) F0 ~ POSITION
+ DIALECT2.POSITION1
+ DIALECT2.POSITION2
+ s(POINT, by = POSITION)
+ s(POINT, by = DIALECT2.POSITION1)
+ s(POINT, by = DIALECT2.POSITION2)
+ s(subj, bs = "re") + s(syll, bs = "re")

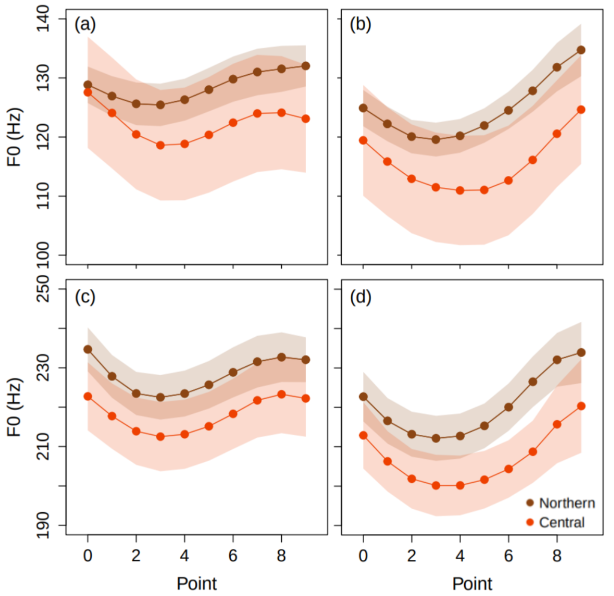

Figure 5: Ten-point pitch extractions in the two dialects for males (a-b) and females (c-d). The first column is the prefinal position and the second is the final. Shaded areas represent standard error.

Table 3 shows the smooth terms of the GAMM results of the difference smooths, which directly compared the tonal excursions of the two dialects while disregarding differences in pitch register. Results showed that except for the prefinal position in females, central speakers differed consistently from their northern counterparts in terms of their tonal contours. Fig. 6 shows the difference smooths predicted by GAMM, which generally started high in the beginning, and gradually declined toward the end. Since the reference level of DIALECT was set at the northern variety, this suggests that compared to their northern counterparts, central speakers tended to have a steeper fall in the beginning, resulting in a large positive difference that gradually diminished, followed by a shallower rise, resulting in a negative difference, the magnitude of which gradually increased.

4. CONCLUSION

Results in this study were interesting yet surprising, as some of the previously assumed patterns were blatantly disconfirmed (cf. [6]). Although rising was indeed found to be fairly common for Tone 1 in the central dialect, as was predicted from studies on other tones [8, 9], it was not exclusive to the variety, nor was it to the final position. This implies the cues that formed the impression of the general public are definitely not as straightforward as one had previously assumed. Based on the results of this study, both dialects showed a preference for a rising Tone 1, and had a comparable degree of rising realisation and a similar extent of the rise. However, the central dialect consistently showed a steeper fall before the rise, which potentially created a stronger contrast to the following rise and thus a percept of a more prominent rise. While the initial falling portion is usually not considered phonologically relevant in a rising contour, perception studies have shown that it could still effectively contribute to the percept of a rising tone [10]. Therefore, we suspect that the steepness of the initial fall might play a key role in forming the impression of the general public for the central dialect. This would merit further studies.

Table 3: Smooth terms of GAMM fitted to male and female F0 contour shape data using the first position from the northern dialect as the reference level.'*' $p < .05$, '**' $p < .01$, '***' $p < .001$.

Smooth terms	**edf**	**Ref.df**	***F* value**	
		Male		
Position1	4.29	5.29	15.38	***
Position2	4.67	5.75	35.23	***
Dialect2:Position1	1.30	1.53	7.19	**
Dialect2:Position2	2.13	2.65	5.30	**
Subject	9.99	12.00	945.71	
Syllable	1.26	2.00	23.39	
		Female		
Position1	5.10	6.22	24.31	***
Position2	5.37	6.54	76.66	***
Dialect2:Position1	1.00	1.00	0.02	
Dialect2:Position2	1.00	1.00	21.98	***
Subject	10.98	13.00	547.39	
Syllable	1.50	2.00	14.49	

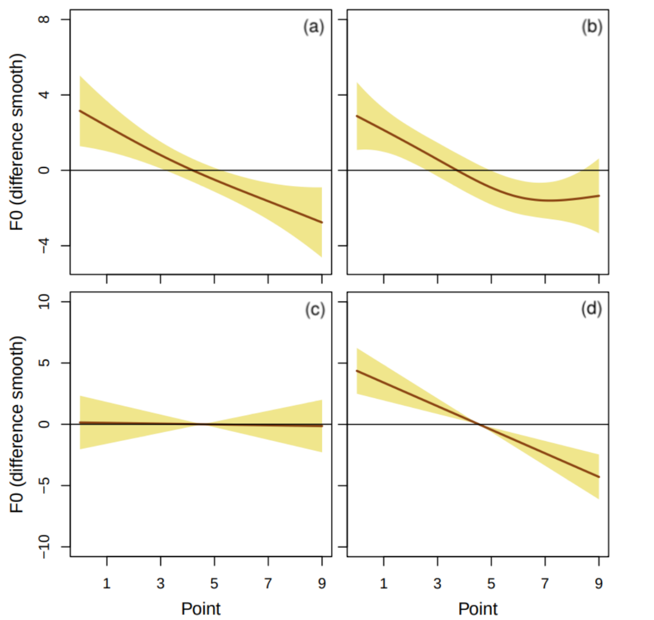

Figure 6: Difference smooths predicted by GAMM for males (a-b) and females (c-d). The first column is the prefinal position and the second is the final. Shaded areas represent two standard errors.

7. REFERENCES

[1] Fon, J., Chiang, W.-Y. 1999. What does Chao have to say about tones? --A case study of Taiwan Mandarin. *Journal of Chinese Linguistics*, 27, 15–37.

[2] Fon, J. 2020. The phonetic realizations of the Mandarin phoneme inventory: The canonical and the variants. In: Liu, H.-M., Tsao, F.-M., Li, P. (eds), *Speech Perception, Production and Acquisition - Multidisciplinary approaches in Chinese languages*. S Springer, 11–36.

[3] Huang, S. 1993. *Language, society, and ethnic identity: A study on language sociology in Taiwan.* Crane.

[4] Kubler, C. C. 1985a. *The Development of Mandarin in Taiwan: A case study of language contact*. Student Book.

[5] Kubler, C. C. 1985b. The influence of Southern Min on the Mandarin of Taiwan. *Anthropological Linguistics*, 27, 156–176.

[6] Khoo, H. 2020. A preliminary study of the tonal features of central Taiwan Mandarin. *Taiwan Journal of Linguistics*, 18, 115–157.

[7] Huang, Y.-H., Fon, J. 2011. Investigating the effect of Min on dialectal variations of Mandarin tonal realization. *Proc. 17th ICPhS Hong Kong*, 918–921.

[8] Fu, J.-W. 1999. Chinese tonal variation and social network-a case study in Tantzu Junior High School Taichung, Taiwan. MA thesis. Providence University.

[9] Wu, S.-J. 2003. A sociolinguistic study of Chinese tonal variation in Puli, Nantou, Taiwan. MA Thesis. Providence University.

[10] Fon, J., Chiang, W.-Y., Cheung, H. 2004. Production and perception of two dipping tones (T2 and T3) in Taiwan Mandarin. *Journal of Chinese Linguistics*, 32, 249–280.

[11] Tsao, F.-F. 2000. Taiwanized Japanese and Taiwan Mandarin--Two Case Studies of Language Contact during the Past Hundred Years in Taiwan. *Hanxue Yanjiu (Chinese Studies)*, 18, 273–297.

[12] Boersma, P., Weenink, D. 2009. Praat: Doing phonetics by computer (Version 5.1) [Computer software]. http://www.praat.org/

[13] Bates, D., Maechler, M., Bolker, B., Walker, S., & Others. (2014). lme4: Linear mixed-effects models using Eigen and S4. *R Package Version*, 1, 1–23.

[14] R Core Team. 2021. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

[15] Barr, D. J., Levy, R., Scheepers, C., Tily, H. J. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68, 255–278.

[16] Wood, S. N. 2017. *Generalized Additive Models: An Introduction with R* (2nd ed.). Chapman and Hall/CRC Press.

[17] Wood, S. N. 2011. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 73, 3–36.