




ARTICLE

When a rise is not only a rise: An acoustic analysis of the impressionistic distinction between northern and central Taiwan Mandarin using Tone 1 as an example

Janice Fon^{1,*}  and Yu-Ying Chuang²

¹Graduate Institute of Linguistics, National Taiwan University, Taiwan and ²Department of Taiwan Culture, Languages and Literature, National Taiwan Normal University, Taiwan

*Corresponding author. Email: jfon@ntu.edu.tw

(Received 18 May 2023; revised 4 March 2024; accepted 26 March 2024)

Abstract

This study looked at the realization of the high-level Tone 1 in Taiwan Mandarin to examine a public impression of the central dialect, which is said to have a tendency to end with a rise. Fifty-three Mandarin native speakers (27 northern and 26 central) were recruited. Half performed a reading task and half a word-guessing task on 24 disyllabic words with Tone 1 embedded. Results showed rising realizations were the most prominent for the tone, regardless of dialect, gender, genre, and syllable position, but were more prevalent among females than males, and more common and enlarged in the final than the non-final position. Dialectal differences were twofold and mainly lay in the acoustic realization. Central speakers showed both a lower pitch register and a steeper declination than their northern counterparts, and central females also demonstrated an upstep in the final position of the word-guessing task, which completely annihilated the effect of the downtrend. This implies the impressionistic tendency to end high indeed exists in the Tone 1 of the central variety, but its percept is not based on rising realizations alone. Instead, it stands out as a dialectal feature via an enlargement of the rise in the foreground against a disruption of the downward trend in the background. The female lead in the realization suggests the rising Tone 1 does not come with a negative connotation. Perceptual tolerance for the variant likely stemmed from a long-standing free variation between high-level and high-rise for the tone.

Keywords: Tone 1 realization; Taiwan Mandarin; dialectal variation; rising tone

1 Introduction

Taiwan Mandarin is the official language of Taiwan, a small island country in Southeast Asia. It is vibrant with phonetic variations, mostly due to constant contact with Min (Kubler 1985a, 1985b), a former *lingua franca* and currently a major substrate language with which over 70% of the population show various degrees of familiarity (Huang 1993). Robust variations documented so far include mainly segmental changes, such as deretroflexion of retroflexes (e.g., Kubler 1985a, 1985b), hypercorrection of dental sibilants (e.g., Chung

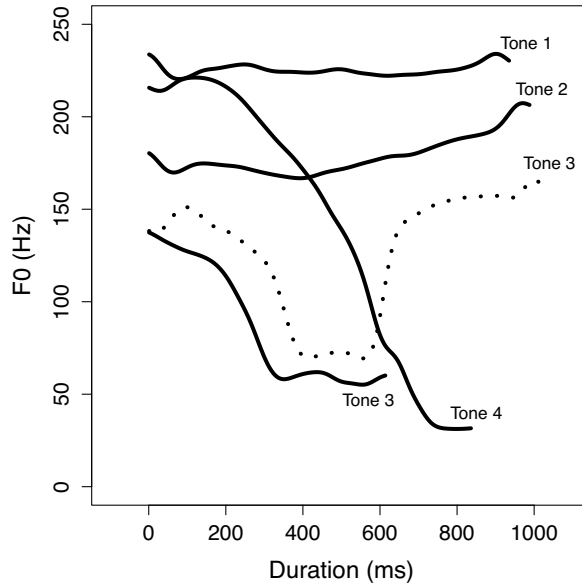


Figure 1. Realization of the four tones in Taiwan Mandarin. Tone 3 has two tonal realizations, a low fall (solid line) and a low dipping (dotted line).

2006), and merging of syllable-final nasals (e.g., Fon et al. 2011), all of which stemmed from contact with Min.

In contrast, even though Mandarin is tonal, and includes a high-level Tone 1, a mid-rising Tone 2, a low-falling/dipping Tone 3, and a high-falling Tone 4 (Chao 1968), as shown in Figure 1, studies on tonal variations have been rare.¹ One exception is the tonal variation in the central dialect (Khoo 2020). Although relevant research is still scant, it has gained public awareness for some time. Michael Shih, an award-winning local pop singer, issued a song called *Taichung Qiang* ‘Taichung accent’ in 2011 (Shih 2011).^{2,3} A Google search on the keyword *Taichung Qiang* as of April 12, 2023 returned 20,100 results (Figure 2).⁴ This implies that the accent should have already been fairly noticeable to Taiwan listeners well before 2011. Khoo (2020) claimed that the term first appeared in the mass media around the 1990s and has become more stable after 2003.

¹ Like its mainland counterpart spoken in China, Tone 3 has the same two allophonic realizations in Taiwan Mandarin, a low fall and a low dipping, but their distributions are somewhat different. Unlike the mainland variety, in which the low fall is pre-final and the low dipping is final (Chao 1968), in Taiwan Mandarin, the low fall has become the default realization regardless of position while the low dipping is mainly reserved for emphatic purposes (Fon & Chiang 1999; Fon, Chiang, & Cheung 2004; Kubler 1985b, 1985a; Tsao 2000).

² Hanyu Romanization is adopted throughout this paper for Mandarin words not otherwise conventionalized. Tonal categories appear as post-syllabic superscript numbers when necessary (see below).

³ Taichung is the largest metropolis in central Taiwan, and thus the term *Taichung Qiang* ‘Taichung accent’ is used by the general public to refer to the variety spoken in central Taiwan, although speakers of this accent do not necessarily come from Taichung only (Khoo 2020).

⁴ A search using keywords of both *Taichung Qiang* ‘Taichung accent’ and Michael Shih returned only 2330 results. This implies that the number of hits for the sole keyword of *Taichung Qiang* could not be all due to searches for the name of the song.

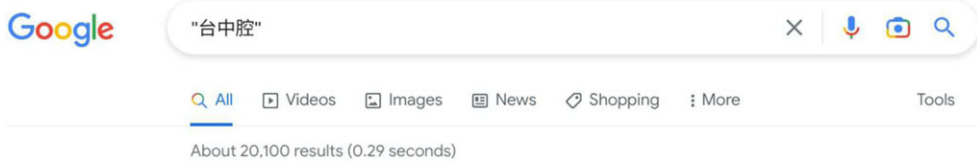


Figure 2. A google search on *Taichung Qiang* “Taichung accent” returned 20,100 results.

1.1 Previous studies

Previous literature showed that there are at least two tonal features associated with the central variety. First, the central dialect favors a lower pitch register and a narrower pitch range compared to the northern standard dialect, and the effect is especially prominent on high targets, including Tone 1, Tone 2, and the beginning of Tone 4 (Huang & Fon 2011; Khoo 2020).⁵ Huang & Fon (2011) suspected that higher Min proficiency might be an underlying cause for lowering of the pitch register, and Khoo (2020) also argued that pitch-range narrowing stemmed from a negative transfer from the local Min variety, which realizes its Tone 1 as mid-level instead of the usual high-level in the mainstream dialects. Since the majority of the population in Taiwan are familiar with Min to various degrees (Huang 1993), the conjecture on Min shaping the pitch register of Mandarin is not surprising and fairly plausible.⁶ Comparing Mandarin in Taiwan and China, both Deng et al. (2006) and Huang & Fon (2011) found that the tonal register of the former is lower than that of the latter, which has little influence from Min. Wu (2009) looked at northern Taiwan Mandarin speakers and found that their tonal register was negatively correlated with their Min proficiency. The higher the Min proficiency, the lower the pitch register. Since northern speakers generally use less Min than their central counterparts (National Statistics R.O.C. 2021), as is shown in Table 1, it is likely that the lower pitch register of the central dialect is also due to Min transfer.

Table 1. Min usage calculated by the percentage of speakers using the language as the primary or the secondary means of daily communication in different regions of Taiwan adapted from the 2020 Population and Housing Census conducted by the Taiwan government (National Statistics R.O.C. 2021).

%	Northern	Central	Southern	Eastern
Primary	18.17	41.40	48.42	18.57
Secondary	61.43	49.52	46.23	47.43

Secondly, the central dialect tends to end high at utterance-final positions (Khoo 2020). Although this feature is often mentioned in the mass media, systematic research is rare.

⁵ Both studies used reference points to determine the pitch range. Huang and Fon (2011) adopted references in a tone-dependent fashion, and included f₀ maximum for Tone 1, initial f₀ maximum, medial f₀ minimum, and final f₀ maximum for Tone 2 and Tone 3, and initial f₀ maximum and final f₀ minimum for Tone 4. On the other hand, Khoo (2020) adopted initial and final f₀ for all four tones.

⁶ Although Min has a larger tonal inventory than Mandarin, there is a tendency for it to adopt a lower and narrower tonal range compared to Mandarin (Chen 2005), contrary to the prediction by the Theory of Adaptive Dispersion (Liljencrants & Lindblom 1971). However, Alexander (2010) found that Cantonese, a Chinese tonal language geographically close to Min, also showed a narrower tonal space and more dispersion from the high tonal target than Mandarin despite its even larger tonal inventory.

To the best of our knowledge, there are only two relevant studies. Fu (1999) examined 16 junior high school classmates from central Taiwan and found that they tend to realize their utterance-final Tone 3s as dipping, instead of the default low fall.⁷ The distribution is gender- and genre-dependent, and interacts with the group status of the speakers. Males generally use the variant more often than females, and predominantly in spontaneous speech. Both core and secondary members of the group consistently adopt the variant in free conversations while peripheral members are not as unanimous, and some avoid it altogether. Female speakers are the exact opposite, and use the variant only in read speech. Peripheral members of the group are more likely to adopt the form than secondary members, while core members completely abstain from using it. The gender-dependent effect of social status could be readily explained by Labov's (2001) gender paradox. Females tend to pursue overt prestige, and their core members thus prefer the standard low-fall over the dipping variant. However, males tend to pursue covert prestige. Therefore, their core members pride themselves on choosing the dipping variant over the low-fall standard instead (Trudgill 1972). The gender-dependent genre effect seems somewhat counterintuitive. One would usually assume variant forms to occur more often in informal spontaneous speech than formal read speech regardless of gender. Based on a subsequent judgment task, Fu (1999) argued that it stems from differential sensitivity for power and solidarity in linguistic forms between the two genders. Females are overall more perceptive of the social connotations attached to the different Tone 3 renditions and tend to dynamically adjust their forms in an interlocutor-dependent manner. When conversing with core members of the group, non-core females tend to switch from the variant to the standard form, resulting in its low usage in conversational speech. However, males seem less aware of the connotations attached and no such effect was observed. Most males tend to use their preferred forms regardless of the social status of their interlocutor.

Wu (2003) conducted a study on the Tone 4 realizations of 54 adult central speakers and found that they are inclined to realize their Tone 4s as high-level, rather than the prescribed high-fall. This is especially true when Tone 4s are at the utterance-final position in spontaneous speech. Tonal environment also plays a role. A Tone 1 + Tone 4 and a Tone 2 + Tone 4 sequence are more likely to elicit a high-level Tone 4 than other tonal combinations. A Tone 4 preceded by a high-level Tone 4 is also more likely to be realized as a high-level Tone 4 itself, as shown in (1). The tonal targets followed Shih (1988). Wu (2003) thus attributed the variant to be a result of preservatory assimilation (Xu 1994, 1997), as the high tonal target of a preceding syllable is carried over to affect the following tone. The variant form is not unanimously adopted by all speakers. Younger speakers in their twenties are almost twice as likely to adopt the form than older speakers in their thirties and forties (46% vs. 24% vs. 28%). This implies that the variant might have originated from the younger generation and have only gradually spread to other age groups. Unlike Fu (1999), there is not an outright gender effect. Both male and female speakers seem to have adopted the form in a comparable fashion. However, like Fu (1999), preference for the variant form is modulated by social connections in a gender-dependent manner. Males with a dense local connection are almost four times as likely to adopt the high-level Tone 4 than those with tenuous ties (44% vs. 12%), while females showed the exact opposite. Those with a loose local bond are more likely to adopt the form than those with a tight connection (46% vs. 34%). A subsequent judgment task showed that speakers tend to align the high-falling Tone 4 as a form

⁷ Since Fu (1999) did not provide acoustic measurements, it is unclear how the dipping Tone 3 variant was realized in her study. However, our impressionistic observations suggest that the central Tone 3 variant tends to be perceptually closer to a rising than a dipping tone. In other words, it occupies the mid-to-high rather than the conventional mid-to-low pitch range. However, for convenience's sake, this study followed Fu's (1999) term and used "dipping tone" to refer to this variant.

of power and the high-level variant as a form of solidarity. Therefore, the result could also be explained by the differential preference for overt and covert prestige between the two genders (Trudgill 1972), similar to that in Fu (1999). Interestingly, the two tonal variants, dipping Tone 3 and high-level Tone 4, are not independent of each other. Speakers who use the high-level Tone 4 also tend to realize their Tone 3 as dipping (Wu 2003). To conclude, the two variants might have arisen from a common preference for ending high, which is in line with layman's impression of the dialect. However, it is worth noting that neither Fu (1999) nor Wu (2003) adopted acoustic analyses in their studies, and thus the exact realization of the tones remains unclear.

- (1) Tone 1 + Tone 4: HH + HL → HH + HH
 Tone 2 + Tone 4: LH + HL → LH + HH
 Tone 4 (high-level) + Tone 4: HH + HL → HH + HH

1.2 Aims of the study

To evaluate whether the preference for ending high is specific to only Tone 3 and Tone 4 or common across all four tones, this study extended the inquiry to Tone 1. Of the two remaining tones, Tone 2 is of little help in clarifying this issue since it is already ending high. However, since Tone 1 is realized with lower pitch in the central dialect (Huang & Fon 2011; Khoo 2020), it could be an ideal candidate, as any preference for ending high could be readily detected through realization of a rise.

This study thus has three aims. First of all, we would like to use acoustic measures to see if Tone 1 tends to be realized with a rise in the central variety. Previous studies of Fu (1999) and Wu (2003) 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 standard dialect. Although studies regarding both varieties are existent (e.g. northern: Kubler (1985a, 1985b), Fon & Chiang (1999), Tsao (2000), Fon, Chiang & Cheung (2004); central: Fu (1999), Wu (2003)), few have included both for direct comparison, except for Huang & Fon (2011) and Khoo (2020), which only focused on the tonal range and register, not the tendency for ending high. Therefore, we included both dialects to directly gauge the degree of divergence between the two. Finally, we would like to examine more closely the effects of syllable position and genre. Both Fu (1999) and Wu (2003) showed that the preference to end high is the most prevalent at utterance-final positions, and spontaneous speech is more likely to elicit high-ending variants than read speech. This is not surprising, since the longer syllable duration allowed at the utterance-final position due to final lengthening likely provides the longer time needed for the rising excursions of the dipping Tone 3 (Xu & Sun 2002) and the sustained pitch of the high-level Tone 4 (Zee 1978) than the pure falling renditions (see Figure 1). Also, unlike read speech, which usually requires careful and standard pronunciations, spontaneous speech tolerates more performance variations and is usually teemed with informal variants due to compromises among various constraints (Labov 1972; Johnson 2004). However, since Fu (1999) and Wu (2003) adopted unscripted conversation for spontaneous speech, and word lists and passages for read speech, it is intrinsically difficult to simultaneously control for both syllable position and syllable number across the two genres. To more precisely examine the effects, this study adopted a more controlled paradigm in order to determine the exact scope of the tendency to end high for different genres.

2 Method

2.1 Subjects

Fifty-three native Mandarin speakers aged between 18 and 27 were recruited in this study ($\bar{X} = 21.04$, $SD = 1.86$). Approximately half were from northern Taiwan ($N = 27$), and half from central. The two regions were comparable in age ($t(51) = -0.87$, ns). All subjects were born and raised in their respective areas and had not lived outside the region for more than six months before they were 18 years old. Subjects were randomly assigned to one of the two groups, the scripted speech group and the unscripted speech group (see Section 2.4). Each group was about equally divided in gender. Please see Table 2 for the distribution.

Table 2. Number of subjects in each group

	Scripted		Unscripted		Total
	Male	Female	Male	Female	
Northern	6	7	7	7	27
Central	6	6	8	6	26
Total	12	13	15	13	53

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 each 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. The scripted speech group included an additional set of 72 disyllabic fillers that did not contain any of the target syllables. The set was downsized to 24 for the unscripted speech group. For the full list of stimuli, please see the Appendix.

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 subsequently down-sampled to 22050 kHz using Praat 6.1 (Boersma & Weenink 2009).

2.4 Procedure

Recording was carried out in a quiet room. For the scripted speech group, subjects performed a word-reading task by reading words printed on index cards in a clear and natural fashion, one word per card. For the unscripted speech group, subjects underwent a word-guessing task and guessed words based on oral hints provided. For example, hints to *ban*¹*jia*¹ “to move (to a new residence)” included the cue phrase “to change residence.” All hints were predetermined and did not contain any syllables in the target words. Subjects were encouraged to guess the answers as quickly as possible. Both groups of subjects were randomly assigned to one of the three semi-randomization orders specific to each task. Words containing the same target syllable were hand-adjusted to avoid juxtaposition.

Although both tasks had the stimuli as the intended targets, subjects' level of pronunciation awareness was different. In the read speech task, since the words were presented in a written form, subjects were more likely to focus on the pronunciation, and tended to adopt renditions deemed appropriate for formal use. On the other hand, in the word-guessing task, subjects' attention was directed to the meaning instead, and were less likely to focus on the pronunciation. In fact, many subjects were so engrossed in the task that they showed many performance features common to spontaneous speech, such as silent and filled pauses, repetitions, and self-corrections. In other words, they were not trying to be "careful" about their speech at all, and were inclined to use a casual speech style. By adopting the two tasks, we hoped to elicit the target stimuli that were matched in syllable number, syllable position, and focus placement, but differed in formality, so that it would be possible to study the distribution of rising Tone 1s that is exclusively modulated by speech style. If rising Tone 1 is considered to be a rendition of solidarity and is associated with a negative connotation, then one would expect it to surface more in the word-guessing game. On the other hand, if rising Tone 1 does not come with a negative connotation, and could also act as a form of power, then its realization rates should be approximately the same across the two tasks.

2.5 Analyses

The voiced portions of the stimuli were hand-labeled using Praat (Boersma & Weenink 2009), 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, whether level, rising, or falling, were hand-labeled based on independent perceptual judgments on the pitch excursions by three phonetically trained native speaker judges, two of whom were the authors.⁸ A tone was categorized as dynamic or level when it was perceived as such, and did not have to bear any resemblance to existing tonal categories in standard Taiwan Mandarin.

To facilitate comparisons of tonal contours across different tokens, two sets of pitch information were extracted for further analyses via Praat scripts. The first set included tone-dependent reference points from each token, following Huang and Fon (2011) and Khoo (2020). For a level contour, it is the maximum pitch. For a rising contour, it is the initial minimum and final maximum of the rising portion, while for a falling contour, it is the initial maximum and final minimum of the falling portion. The reference points of a contour tone may not be in the absolute initial or final position because the acoustic realization might not be in the form of a pure rise or a pure fall (please see the Results, Section 3). The second set extracted ten pitch points at equal time intervals from each token, so that analyses could be performed on time-normalized tonal contours.

The common practice of pitch normalization through semitone conversion was not adopted in this study (cf. Fon & Chiang 1999, Khoo 2020, among others) because we would like to preserve the critical dialectal distinction in absolute pitch height between the northern and the central variety (Huang & Fon 2011; Khoo 2020), which would be difficult to capture after conversion. Instead, pitch normalization was done statistically through the by-subject random intercept of mixed models.

⁸ Since not all acoustic undulations are phonetically and phonologically meaningful, perceptual judgments were adopted in this study for tonal contour differentiation. However, we are aware that perception is easily affected by various factors, such as loudness, duration, and spectral energy distribution. Therefore, there might be cases in which perception and acoustics do not exactly match. We hope this drawback could be largely mitigated by incorporating judgments from multiple judges. However, we do acknowledge that our analyses are limited to tonal shapes that could be discerned by most trained ears only.

3 Results

3.1 Realization of Tone 1

In total, 24 (stimuli) \times 53 (subjects) = 1272 tokens of stimuli were collected, among which 880 were perceived as rising contours by at least two judges, accounting for 69% of the data (Table 3). In other words, the rising realization was the most prevalent for both dialects, and all speakers adopted it to various extents. Figure 3 shows the average pitch excursion for the perceived level and rising contours in both male and female speakers. The two pitch contours occupied similar pitch registers, with the rising contour being slightly lower than the level one. They also had similar initial trajectories, likely due to the coarticulatory interaction between the onset consonant and the following rhyme. Major differences between the two mainly lay in the second half of the tone, in which the perceptually rising contour showed an audible rise, while the level one did not.

Figure 4 shows the mean duration of the two perceived realizations. Syllable position showed a large effect. The second syllable was on average 55–75 ms longer than the first. Tonal realization had a smaller effect. Rising realizations were generally 5–25 ms longer

Table 3. 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. The “undecided” category refers to tonal contours that did not reach a majority vote.

	Level	Rising	Falling	Undecided	Total
<u>Scripted</u>					
Northern	48/12	106/142	1/1	1/1	156/156
Central	60/18	71/120	10/5	3/1	144/144
<u>Unscripted</u>					
Northern	71/19	79/140	13/6	5/3	168/168
Central	71/19	81/141	13/7	3/1	168/168
Total	250/68	337/543	37/19	12/6	1272

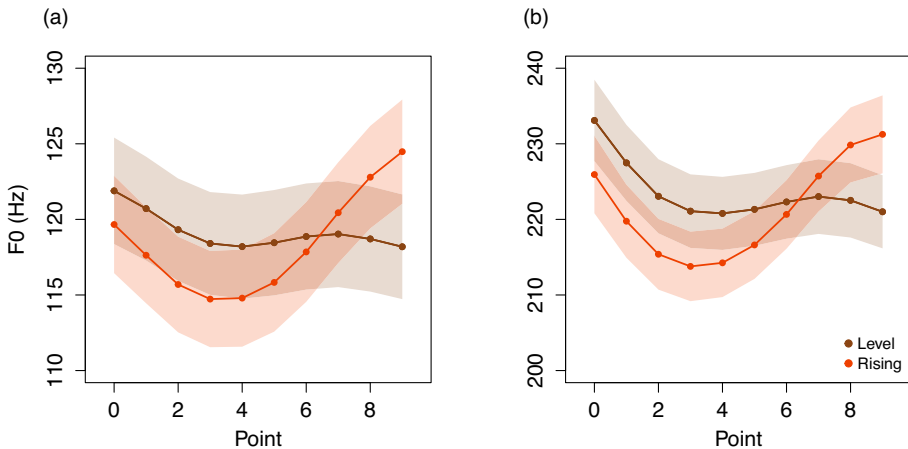


Figure 3. Time-normalized mean pitch excursions of perceived level and rising contours in (a) male and (b) female speakers. Shaded areas represent standard error.

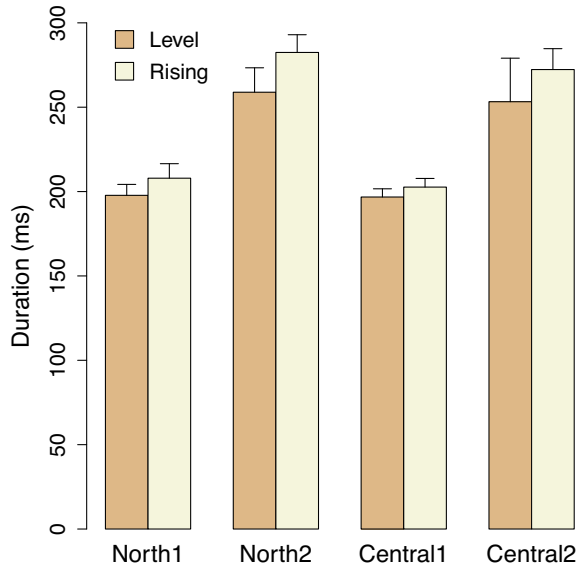


Figure 4. Mean duration of perceived level and rising contours in the two dialects. Error bars represent standard error. “North1” and “North2” refer to the first and second positions of the northern dialect, and “Central1” and “Central2” refer to the first and second positions of the central dialect.

Table 4. Fixed effects of the linear mixed model for duration using the first position of the level realization in the northern dialect as the reference. ** $p < .05$, *** $p < .01$, **** $p < .001$.

	Estimate	S.E.	t value
(Intercept)	196.47	6.08	32.32***
DIALECT2	-0.22	8.36	-0.03
POSITION2	80.88	6.99	6.23***
SHAPE2	11.29	5.64	2.00*
DIALECT2:POSITION2	-29.46	18.04	-1.63
DIALECT2:SHAPE2	-4.51	8.06	-0.56
POSITION2:SHAPE2	-6.13	8.20	-0.75
DIALECT2:POSITION2:SHAPE2	24.01	11.23	2.14*

than level ones. A linear mixed effects analysis using the `lme4` package (Bates et al. 2015) in R (R Core Team 2021) was performed. Fixed effects included DIALECT, POSITION, and SHAPE, along with their interaction terms, and random effects included by-subject and by-item intercepts, and by-subject random slopes for POSITION and SHAPE, as shown in (2). Significance was calculated using normal approximation (Barr et al. 2013). Results confirmed our observations (Table 4). Both POSITION ($p < .001$) and SHAPE ($p < .05$) had a main effect. In addition, there was a significant three-way interaction ($p < .05$). This was because the central dialect had a much larger SHAPE effect in the second position than the first, compared to the northern dialect. The difference between level and rising contours was only about 6 ms in the first position for central speakers, but increased more than three

Table 5. Fixed effects of the mixed effects logistic regression on target CVN syllables using the first position in scripted speech in northern females as the reference. ** $p < .05$, *** $p < .01$, **** $p < .001$.

	Estimate	S.E.	z value
(Intercept)	1.13	0.36	3.13**
DIALECT2	-0.33	0.31	-1.06
GENDER2	-0.88	0.32	-2.77**
GENRE2	-0.35	0.32	-1.12
POSITION2	2.37	0.37	6.40***

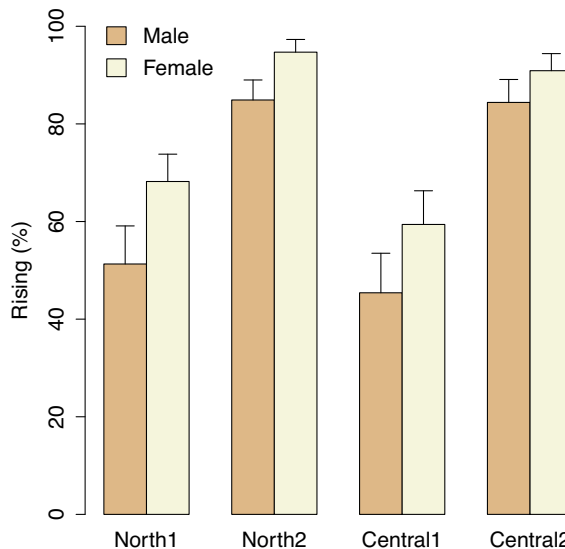


Figure 5. Rising percentages regarding POSITION and GENDER for the CVN stimuli in the two dialects.

times to about 19 ms in the second. On the other hand, although the difference was a slightly larger 10 ms in the first position for northern speakers, it only increased a little more than twofold to 23 ms in the second. However, the two dialects did not seem to differ drastically in their overall duration patterning with regards to SHAPE and POSITION.

(2) DURATION \sim DIALECT * POSITION * SHAPE + (1+POSITION+SHAPE|subject) + (1|item)

To study how the distribution of rising contours was affected by various factors, a mixed effects logistic regression was performed on the tonal shapes. Fixed factors included DIALECT, GENDER, GENRE, and POSITION, while random effects included by-subject and by-item intercepts, and by-subject random slopes for POSITION, as shown in (3). Results indicated that only GENDER and POSITION were significant (Table 5). Females were about 5-15% more likely to have rising realizations than males ($p < .01$), and the second syllable was about 25-40% more likely to be realized as rising contours than the first ($p < .001$). No effects involving DIALECT or GENRE were observed (Figure 5).

(3) SHAPE \sim DIALECT + GENDER + GENRE + POSITION + (1+POSITION|subject) + (1|item)

The lack of a dialectal effect was contrary to expectation (cf. Khoo 2020). Since previous studies did not control for syllable structure but used a variety of syllable types (Fu 1999; Wu 2003), we would like to verify that the current results did not stem from our exclusive adoption of CVN syllables. Therefore, the tonal contours of the pairing syllables that are Tone 1 and of a non-CVN structure were examined. There were three of this type, the *jiá*¹ “home” in *ban¹jiá¹* “to move (to a new residence)”, the *dié*¹ “father” in *gan¹dié¹* “godfather”, and the *zhu*¹ “pig” in *zhu¹gan¹* “pork liver”. There were in total 3 (pairing syllables) × 53 (subjects) = 159 tokens, all of which were subjected to shape judgments by the same three native judges using the same criteria. Results showed that although rising contours were not as commonly found in the first position, they are still the predominant realizations in the second position, accounting for 74% of the data (Table 6).

Table 6. Perceived tonal contours of the non-CVN pairing syllables that are Tone 1 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. The “undecided” category refers to tonal contours that did not reach a majority vote.

	Level	Rising	Falling	Undecided	Total
Northern	15/10	3/43	9/1	0/0	27/54
Central	17/13	2/35	7/3	0/1	26/52
Total	32/23	5/78	16/4	0/1	159

To examine the factors underlying the contour realization of the pairing syllables, a mixed effects logistic regression was performed, with fixed factors of DIALECT, GENDER, GENRE and POSITION, and random effects of by-subject intercepts, as shown in (4). The by-item intercepts were excluded because no sufficient cross-item variation could be detected. Results showed that only the two main effects of GENDER and POSITION were significant. The effects of DIALECT and GENRE were not significant (Table 7). As shown in Figure 6, females generally were about 15–20% higher than males in their rising realization rates, and the second position was about 60–70% more likely than the first in realizing a Tone 1 as a rising contour. A comparison between Figures 5 and 6 shows that the CVN syllables in the stimuli might have boosted the rising realization rate for the non-final syllables, but Tone 1 was realized as predominantly a rising contour in the final position regardless of syllable structure. More crucially, no dialectal difference was found.

(4) SHAPE ~ DIALECT + GENDER + GENRE + POSITION + TASK + (1|subject)

Table 7. Fixed effects of the mixed effects logistic regression on the non-CVN pairing syllables of Tone 1 using the first position in scripted speech in northern females as the reference. ‘.’ $p < .10$, ‘*’ $p < .05$, ‘**’ $p < .01$, ‘***’ $p < .001$.

	Estimate		
	e	S.E.	z value
(Intercept)	−1.10	0.58	−1.88.
DIALECT2	−0.48	0.46	−1.06
GENDER2	−1.14	0.50	−2.30*
GENRE2	−0.35	0.46	−0.76
POSITION2	3.42	0.73	4.71***

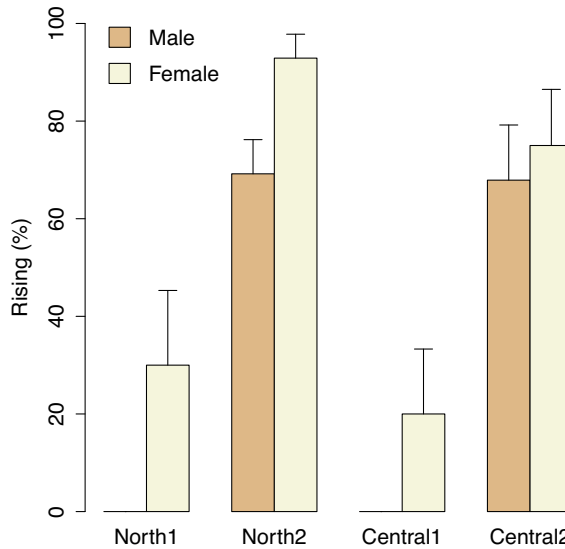


Figure 6. Rising percentages regarding POSITION and GENDER for the non-CVN pairing syllables in the two dialects. “North1” and “North2” refer to the first and second positions of the northern dialect, and “Central1” and “Central2” refer to the first and second positions of the central dialect.

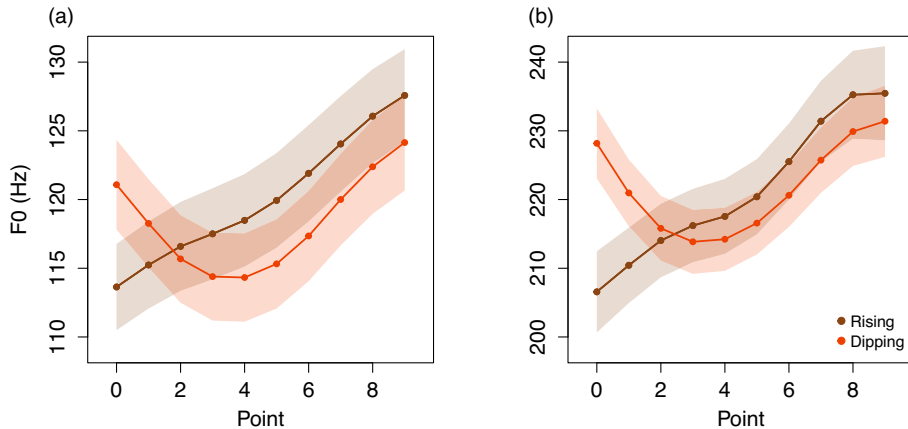
The absence of dialectal difference is rather intriguing. If one assumes that native speakers’ impression of the dialectal divide is perceptually valid (Khoo 2020), then there ought to be some measures out there that are reflective of such perception, although it might not be as straightforward as one had previously conceived. Therefore, we turned to the phonetic realizations of the rising contour to see if that is where the main dialectal variation lies.

3.2 Phonetic realization of rising Tone 1

In Mandarin, there are two common ways to phonetically realize a perceptually rising tone, pure rise and slight dip. Previous studies have found this to be true for the realization of the rising Tone 2 in both the Taiwan (Fon & Chiang 1999) and the Mainland (Shi & Wang 2006) variety, and both varieties unanimously adopted slight dipping realizations more often than pure rising ones (see Figure 1). The two renditions are perceptually discernible to a trained ear, but both are considered to be good realizations of Tone 2 (Fon 2020). No clear rules have been found with regards to the distribution of the two forms, and they seem to be in free variation with each other. It is thus interesting to find that similar realizations were observed among our Tone 1 tokens that were perceived as rising contours. Both the pure rising and the slight dipping renditions were found based on the f_0 tracings using Praat. As shown in Figure 7, the pure rising contour started somewhat lower than the slight dipping rendition, but occupied a slightly higher register during the second half of the tone. Table 8 displays the distribution of the two rising realizations across the two dialects. As evident from the table, most of the perceptually rising Tone 1s were indeed realized as dipping acoustically, and the tendency was higher for the central than the northern variety (91% vs. 81%). A Pearson’s chi-square test with Yates’ continuity correction showed that the difference was significant ($\chi^2(1) = 17.10, p < .0001$).

Table 8. Distribution of the two acoustic renditions of rising Tone 1s in the target stimuli.

	Rising	Dipping	Total
Northern	87	380	467
Central	36	377	413
Total	123	757	880

**Figure 7.** Two renditions of time-normalized mean pitch excursions of Tone 1 tokens that were perceived as rising contours in (a) male and (b) female speakers. Shaded areas represent standard error.

As mentioned previously, two sets of analyses were conducted to explore the potential features that constitute the native speakers' impression of the dialectal distinction. The first incorporated both renditions of a perceptually rising tone by examining the initial minimum and final maximum pitch of the rising portion. This would be the beginning and end pitch of the rising rendition, and the medial lowest and final highest pitch for the dipping rendition. The second complements the first by looking at the overall tonal contours. Since the majority of the rising tones were realized as slightly dipping, only this rendition was examined in the whole-contour analysis.

Figure 8 shows the distribution of the mean pitch height across dialect and gender. Mean pitch was calculated by averaging the ten pitch points extracted from the whole tonal contour. In general, the central dialect tended to be slightly lower in pitch than the northern dialect, regardless of gender, which was in line with previous studies (Huang & Fon 2011; Khoo 2020). However, one central female speaker, CGF, was found to be an outlier. She had extraordinarily high pitch, with an average pitch height of 301.50 Hz ($s = 30.27$), which was in stark contrast to the average pitch of all the other female speakers ($\bar{X} = 218.11$ Hz), and was more than 3 standard deviations away from the overall mean female pitch. She also had the largest pitch range (174 Hz) and highest pitch floor (248 Hz) of all speakers (mean range = 78 Hz and mean pitch floor = 184 Hz for all the other females). Therefore, we suspected that she was an influential outlier to the overall pattern of the data. Two linear mixed models, one with CGF and the other without, were performed to confirm this. Fixed factors included DIALECT and GENDER, and random effects included by-subject and by-item intercepts, as is shown in (5). Results indeed showed that both the effects of DIALECT ($p < .05$) and GENDER ($p < .0001$) were significant when CGF was excluded, while only the GENDER effect was significant ($p < .0001$) when CGF was included. The DIALECT effect was

Table 9. Fixed effects of the two linear mixed effects models on 10-point pitch extractions of rising Tone 1 tokens using the northern females as the reference. ‘*’ $p < .05$, ‘**’ $p < .01$, ‘***’ $p < .001$.

	Estimate	S.E.	z value
<u>with CGF</u>			
(Intercept)	224.31	4.93	45.47***
DIALECT2	-5.74	5.67	-1.01
GENDER2	-102.51	5.67	-18.06***
<u>without CGF</u>			
(Intercept)	222.55	4.08	54.57***
DIALECT2	-9.38	4.69	-2.00*
GENDER2	-98.92	4.69	-21.08***

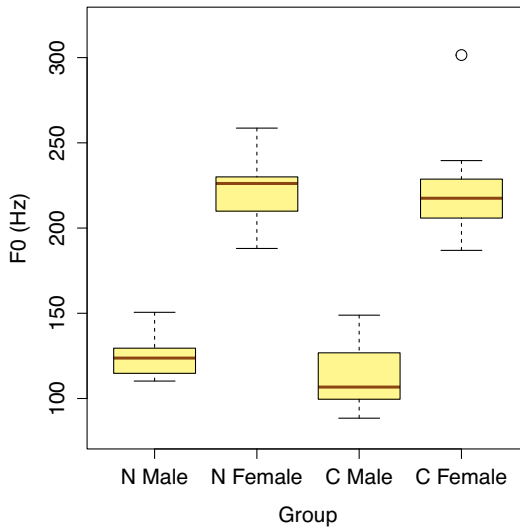


Figure 8. The distribution of mean pitch across DIALECT and GENDER. ‘N’: northern; ‘C’: central. The outlier was Speaker CGF.

not significant anymore (Table 9). As the data of CGF were far away from the others and did not follow the general pitch trend of the central dialect in terms of pitch register, they were considered as atypical and were excluded from further analyses.

$$(5) F_0 \sim \text{DIALECT} + \text{GENDER} + (1|\text{subject}) + (1|\text{item})$$

3.2.1 Reference points of rising Tone 1

Figure 9 shows the mean pitch of the beginning and end reference points of the rise. Position seemed to play a role in the realization of the rising Tone 1. The second syllable tended to be lower than the first. Males and females had an average lowering of 4.75 Hz and 10.25 Hz, respectively. The second position also had a larger rise. Males increased from an average rise of 8.75 Hz in the first position, to 14.5 Hz in the second, while females rose

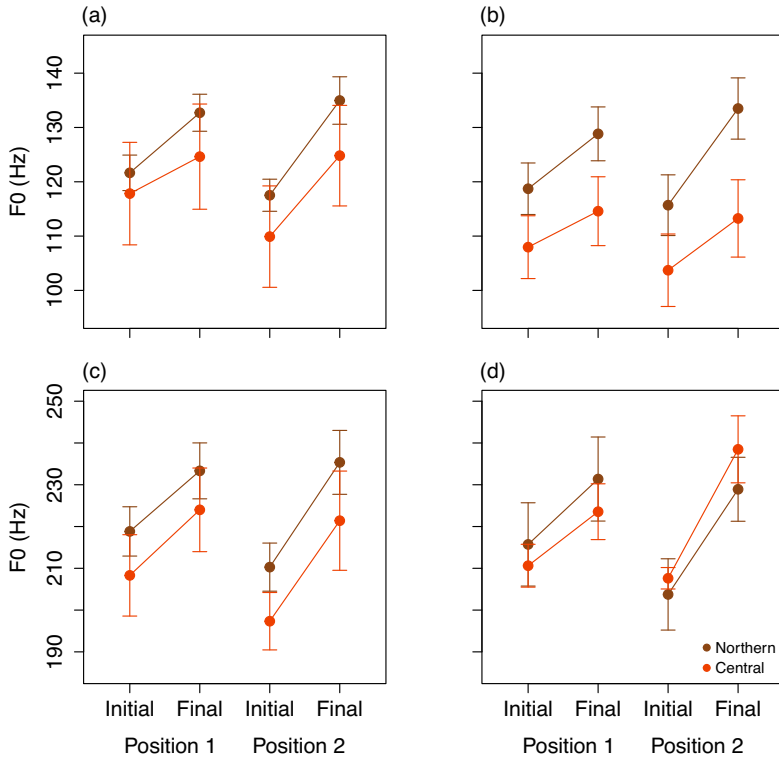


Figure 9. Mean pitch of the beginning and end points of the rise across DIALECT for (a) scripted and (b) unscripted speech for males and (c) scripted and (d) unscripted speech for females. Error bars represent standard error.

from 15.5 Hz to 26.25 Hz. However, the trend was not unanimous across different speaker groups and seemed both gender- and dialect-dependent. Two separate linear mixed effects analyses were performed on males and females to confirm this. The two genders were analyzed separately so as to facilitate clearer observations on the potential interaction between GENDER and other factors. As shown in (6), the model included raw pitch extraction values as the dependent variable, and DIALECT, GENRE, POSITION, and REF_POINT, along with their interaction terms, as the fixed effects. Random effects included by-subject and by-item intercepts, as well as the by-subject random slopes for REF_POINT. The by-item random slopes were not included because inclusion of such would result in non-convergence of the model.

$$(6) F0 \sim \text{DIALECT} * \text{GENRE} * \text{POSITION} * \text{REF_POINT} + (1 + \text{REF_POINT} | \text{subject}) + (1 | \text{item})$$

Table 10 shows the fixed effects for males. The main effect of REF_POINT was significant ($p < .0001$). This was expected, since reference points were extracted from the rising portion of a tone and final points were naturally higher than initial ones. A two-way interaction of POSITION \times REF_POINT, and the four-way interaction were also significant. To examine how DIALECT was involved in the higher-order interactions, the data was subset by DIALECT using the model in (7). Results showed a dialectal split. For northern males, there was no effect involving GENRE. Only the main effect of REF_POINT ($p < .0001$) and the interaction effect of POSITION and REF_POINT ($p < .01$) were significant. The main effect of POSITION was not significant (Table 11a). This suggested that northern males did not show

Table 10. Fixed effects of the linear mixed model for males using the beginning point of the first position in northern scripted speech as the reference. ** $p < .05$, *** $p < .01$, **** $p < .001$.

	Estimate	S.E.	t value
(Intercept)	120.63	6.31	19.11****
DIALECT2	-4.89	8.87	-0.55
GENRE2	-2.28	8.53	-0.27
POSITION2	-2.78	1.90	-1.46
REF_POINT2	11.00	2.07	5.32****
DIALECT2:GENRE2	-5.48	11.92	-0.46
DIALECT2:POSITION2	-3.54	2.15	-1.64
DIALECT2:REF_POINT2	-4.40	3.10	-1.42
GENRE2:POSITION2	0.08	2.04	0.04
GENRE2:REF_POINT2	-1.61	2.93	-0.55
POSITION2:REF_POINT2	6.54	2.00	3.28**
DIALECT2:GENRE2:POSITION2	2.00	2.93	0.68
DIALECT2:GENRE2:REF_POINT2	1.52	4.20	0.36
DIALECT2:POSITION2:REF_POINT2	2.53	3.01	0.84
GENRE2:POSITION2:REF_POINT2	2.49	2.85	0.87
DIALECT2:GENRE2:POSITION2:REF_POINT2	-8.48	4.11	-2.07*

much declination across the two positions (≈ 3 -4 Hz), but the degree of the rise for the second position was much more than the first (17 Hz vs. 10.5 Hz).

$$(7) F_0 \sim \text{GENRE} * \text{POSITION} * \text{REF_POINT} + (1 + \text{REF_POINT} | \text{subject}) + (1 | \text{item})$$

In contrast, the main effect of POSITION was significant among central males ($p < .001$). There was a prominent downward trend from the first syllable to the second (≈ 4 -8 Hz). The three-way interaction was also significant ($p < .05$), which implied a differential interaction effect of POSITION and REF_POINT for the two genres. To confirm, separate linear mixed models were built, as shown in (8). Results showed that the interaction effect was only significant in scripted ($p < .0001$) but not unscripted speech (Tables 11b and 11c). In other words, the second syllable had a significantly larger rise than the first only in scripted speech (15 Hz vs. 7 Hz), while no statistical difference between the two positions was found in unscripted speech (9 Hz vs. 7 Hz).

$$(8) F_0 \sim \text{POSITION} * \text{REF_POINT} + (1 + \text{POSITION} | \text{subject}) + (1 | \text{item})$$

Table 12 shows the fixed effects for females. The main effect of REF_POINT and the interaction effect of POSITION and REF_POINT were significant. As observed from Figures 9c and 9d, the degree of the rise was much larger in the second position than the first (26.25 Hz vs. 15.5 Hz). There were also a main effect of POSITION and a three-way interaction of DIALECT \times GENRE \times POSITION. This suggested that the effect of POSITION was dialect- and genre-dependent. To confirm this, separate models were run on the two dialects using the model in (9). Results showed that the interaction was mainly due to central speakers, as both their main and the interaction effects involving POSITION were significant (Table 13b). One could clearly see where the interaction lay from Figures 9c and 9d. For scripted speech, there was a prominent downward trend from the first to the second syllable (≈ 11 Hz). However, for unscripted speech, declination was almost completely annihilated. Although the initial

Table 11. Fixed effects of the linear mixed models for (a) northern male speech using the beginning point of the first position in scripted speech as the reference, (b) central male scripted speech and (c) central male unscripted speech using the beginning point of the first position as the reference. ** $p < .05$, *** $p < .01$, **** $p < .001$.

	Estimate	S.E.	t value
(a) Northern			
(Intercept)	120.93	4.54	26.64***
GENRE2	-2.53	6.13	-0.41
POSITION2	-3.21	1.97	-1.63
REF_POINT2	10.91	2.48	4.40***
GENRE2:POSITION2	0.38	2.34	0.16
GENRE2:REF_POINT2	-1.65	3.50	-0.47
POSITION2:REF_POINT2	6.62	2.29	2.89**
GENRE2:POSITION2:REF_POINT2	2.52	3.27	0.77
(b) Central-scripted			
(Intercept)	116.14	8.41	13.80***
POSITION2	-6.59	2.04	-3.23**
REF_POINT2	6.59	1.94	3.39***
POSITION2:REF_POINT2	9.08	1.95	4.66***
(c) Central-unscripted			
(Intercept)	108.35	6.09	17.80***
POSITION2	-4.70	1.82	-2.59**
REF_POINT2	6.85	1.38	4.96***
POSITION2:REF_POINT2	2.71	1.48	1.83

syllable started out at approximately the same pitch height as its scripted counterpart in the first syllable (211 Hz vs. 208 Hz), the pitch register was markedly raised for the second syllable, resulting in minimal declination (≈ 3 Hz). In other words, for unscripted speech, central females reversed the effect of declination and replaced it with an upstep instead. For northern females, however, none of the effects was significant and neither declination or upstep was observed (Table 13a).

(9) $F_0 \sim \text{GENRE} * \text{POSITION} + (1 + \text{POSITION} | \text{subject}) + (1 + \text{POSITION} | \text{item})$

3.2.2 Whole contours of rising Tone 1

For the whole-contour analyses, only the predominant slight dipping contour was examined, as it constituted the majority. Since the contours were no longer linear, GAMM (Wood 2017) was adopted to better capture the tonal undulation along the time domain. GAMM is often used to model nonlinear effects, and has been applied to phonetic analyses involving measurements that vary across time, including f_0 contours (Kösling et al. 2013; Chuang et al. 2021; Sun & Shih 2021) and movement trajectories of articulators (Wieling et al. 2016; Tomaschek et al. 2018).

In order to characterize contour differences between the two dialects, GAMM separates tonal excursions into tonal height and tonal contours by using parametric coefficients and smooth terms, respectively. To illustrate, Figure 10a shows two tonal contours of Tones A and B, which differ in both f_0 height and shape. Tone A occupies a higher pitch register and has a concave contour while Tone B occupies a lower pitch register and has a convex

Table 12. Fixed effects of the linear mixed model for females using the beginning point of the first position in northern scripted speech as the reference. ** $p < .05$, *** $p < .01$, **** $p < .001$.

	Estimate	S.E.	t value
(Intercept)	218.48	6.57	33.25***
DIALECT2	-9.58	9.47	-1.01
GENRE2	-3.50	9.12	-0.38
POSITION2	-8.21	3.22	-2.55*
REF_POINT2	14.55	3.82	3.81***
DIALECT2:GENRE2	4.78	13.81	0.35
DIALECT2:POSITION2	-3.92	3.61	-1.09
DIALECT2:REF_POINT2	-1.29	5.92	-0.22
GENRE2:POSITION2	-3.02	3.53	-0.86
GENRE2:REF_POINT2	0.97	5.73	0.17
POSITION2:REF_POINT2	10.76	3.21	3.35***
DIALECT2:GENRE2:POSITION2	12.39	5.44	2.28*
DIALECT2:GENRE2:REF_POINT2	-1.43	8.75	-0.16
DIALECT2:POSITION2:REF_POINT2	0.61	5.08	0.12
GENRE2:POSITION2:REF_POINT2	-0.67	4.95	-0.13
DIALECT2:GENRE2:POSITION2:REF_POINT2	7.83	7.62	1.03

Table 13. Fixed effects of the linear mixed model for (a) northern and (b) central females with regards to GENRE and POSITION using the first position in scripted speech as the reference. ** $p < .05$, *** $p < .01$, **** $p < .001$.

	Estimate	S.E.	t value
(a) Northern			
(Intercept)	227.60	7.45	30.54***
GENRE2	-2.88	10.36	-0.28
POSITION2	-3.15	2.58	-1.22
GENRE2:POSITION2	-3.44	2.26	-1.55
(b) Central			
(Intercept)	199.01	6.48	30.73***
GENRE2	-2.12	9.16	-0.23
POSITION2	-7.72	3.62	-2.13*
GENRE2:POSITION2	12.69	3.54	3.59***

contour. The difference between the two curves (i.e., A minus B at each time point) is therefore larger at the ends and diminishes gradually towards the center, as shown in Figure 10b. Since Tone A is consistently higher in pitch than Tone B, there is a nonzero difference between the two across the whole time domain. We could further abstract away the tonal height differences and focus only on contour comparisons by aligning the pitch of the two tones at the tone center. As shown in Figure 10c, the difference contour is shifted downward so that $y = 0$ is tangent to the curve. Through parametric coefficients and smooth

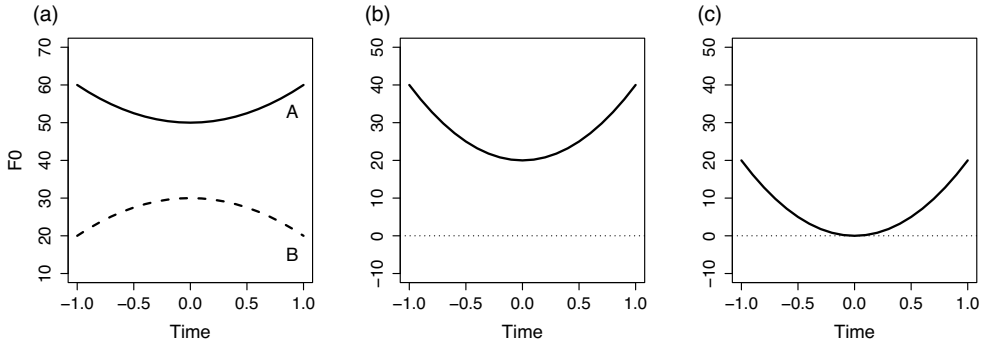


Figure 10. A demonstration of modeling the difference between two contours by using GAMM. The two tones in (a) are different in both f0 height and contour. Their point-wise raw contour differences are shown in (b), and their pure contour differences with the height differences extracted away are shown in (c).

terms, GAMM could thus model tonal height and contour, respectively, and provide direct comparisons on tonal trajectories.

Figure 11 shows the 10-point extractions for the slightly dipping Tone 1 from both northern and central speakers. Separate GAMM models were built on males and females to fit the pitch extraction values using the `bam` function of the `mgcv` package (Wood 2011) in R (R Core Team 2021). We used GAMMs to model height and contour differences of the two dialects directly. As shown in (10), we first specify the modeling of pitch height. The first row asks the model to predict a height estimate for each level of GENRE by POSITION for the northern dialect, which is the reference level. The subsequent four rows request the prediction of height adjustment for the central dialect (DIALECT2) for each of the GENRE.POSITION levels. The second half of the formula specifies the modeling of the tonal contour. Similar to height, the first smooth term requests a contour prediction of each GENRE.POSITION level for the northern dialect. The following four smooths are specified as difference smooths, which model directly the contour differences between the northern and central dialects for each GENRE.POSITION level. Finally the by-subject and by-item random intercepts are added, as indicated in the final row.

```
(10) F0 ~ GENRE.POSITION
      + DIALECT2:GENRE1:POSITION1
      + DIALECT2:GENRE1:POSITION2
      + DIALECT2:GENRE2:POSITION1
      + DIALECT2:GENRE2:POSITION2
      + s(POINT, by = GENRE.POSITION)
      + s(POINT, by = DIALECT2:GENRE1:POSITION1)
      + s(POINT, by = DIALECT2:GENRE1:POSITION2)
      + s(POINT, by = DIALECT2:GENRE2:POSITION1)
      + s(POINT, by = DIALECT2:GENRE2:POSITION2)
      + s(subject, by="re") + s(item, by="re")
```

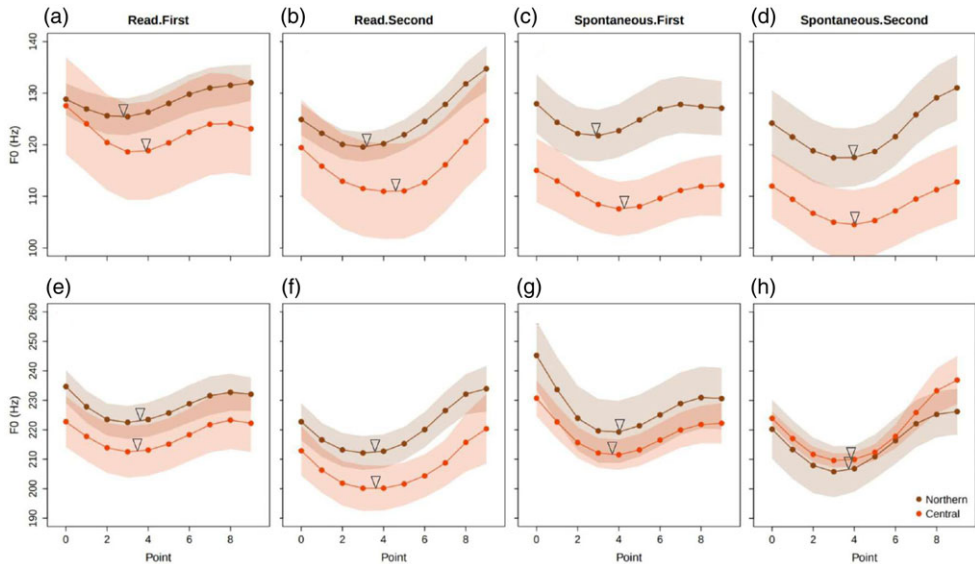


Figure 11. Ten-point pitch extractions in northern and central dialects for male (a–d) and female speakers (e–h). The first and second columns respectively represent the first and second syllables of the scripted speech, while the third and fourth columns respectively represent the first and second syllables of the unscripted speech. Inverted triangles indicate the average divide between the falling and the rising arm. Shaded areas represent standard error.

Table 14 shows the GAMM results for the male speakers. It is interesting to find that dialectal differences mainly lay in smooth terms, not parametric coefficients. In other words, central speakers differed consistently from their northern counterparts in terms of their tonal contours in all four treatment conditions of GENRE by POSITION combinations. As for tonal height, even though there seemed to be a trend for central speakers to occupy a lower pitch register, as is shown in Figures 11a–d, the variability was likely too large for the model to detect significance. Figures 12a–d show the difference smooths predicted by GAMM. For male speakers the difference generally started high in the beginning, and gradually declined toward the end. Since the difference smooths indicate the point-wise adjustment required for the central dialect, this suggests that compared to their northern counterparts, central males generally had 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). The trend was applied across genres and positions.

Table 15 shows the GAMM results for the female speakers. Like males, dialectal differences did not show up as significant in parametric coefficients, even though the central dialect did seem lower in pitch register for all except for the final position in unscripted speech (Figures 11e–h). The variation was again likely too large for the comparison to reach significance. Unlike males, however, there was also not much dialectal difference in tonal contour. Except for the final position in scripted speech, none of the smooth terms reached significance. This implies that the tonal contour for females was to a large extent similar between the two dialects, which could also be observed in the difference smooths predicted by GAMM in Figures 14e, g, and h. In all three cases, the lines hover around $y = 0$, indicating no detectable difference between the contours. For the final position of scripted speech (Figure 12f), the difference smooth was fairly similar to the ones in males. It started high in the beginning, and gradually declined throughout the tone. Since the reference level was also set at the northern variety, this implies that compared to their northern counterparts,

Table 14. Summary of GAMM fitted to male f0 data using the first position in scripted speech from the northern dialect as the reference level. Boldface indicates dialectal differences for tonal height (parametric coefficients) and contour shape (smooth terms). '*' $p < .05$, '**' $p < .01$, '***' $p < .001$.

Parametric coefficients	Estimate	S.E.	t value
(Intercept)	107.79	10.98	9.82***
GENRE1:POSITION2	-3.68	1.36	-2.71***
GENRE2:POSITION1	-6.48	9.24	-0.70
GENRE2:POSITION2	-6.50	9.33	-0.70
DIALECT2:GENRE1:POSITION1	-5.31	6.78	-0.78
DIALECT2:GENRE1:POSITION2	-6.82	6.78	-1.01
DIALECT2:GENRE2:POSITION1	-6.80	6.08	-1.12
DIALECT2:GENRE2:POSITION2	-9.49	6.08	-1.56
Smooth terms	edf	Ref.df	F value
GENRE1:POSITION1	4.47	5.51	16.07***
GENRE1:POSITION2	4.95	6.06	37.79***
GENRE2:POSITION1	4.47	5.52	21.70***
GENRE2:POSITION2	5.36	6.51	39.04***
DIALECT2:GENRE1:POSITION1	1.53	1.88	5.65**
DIALECT2:GENRE1:POSITION2	2.49	3.11	5.94***
DIALECT2:GENRE2:POSITION1	1.00	1.00	5.03*
DIALECT2:GENRE2:POSITION2	2.66	3.30	8.34***
SUBJECT	22.98	24.00	2817.56***
ITEM	21.48	23.00	1386.78***

central females tended to start off with a steeper fall, followed by a shallower rise in this position.

As shown in Figure 11, the fall-to-rise ratio of the tonal contour also demonstrated an interesting pattern. Although the rising proportion was generally larger than the falling one ($t(52) = -8.12, p < .0001$), the fall-to-rise ratio was dialect- and gender-dependent. For males, there was a tendency for the central speakers to have a proportionately longer falling arm than that of their northern counterparts, while for females, the ratios seemed fairly comparable across the two dialects. Separate linear mixed effects analyses were performed on the falling proportion for the two genders to verify the observation, with fixed effects of DIALECT, GENRE, and POSITION, along with their interaction terms, included in the model, as shown in (11). The by-subject and by-intercept were entered as random effects.

$$(11) \text{ FALLING\%} \sim \text{DIALECT} * \text{GENRE} * \text{POSITION} + (1|\text{subject}) + (1|\text{item})$$

For males, results showed that the main effect of DIALECT was indeed significant (Table 16). The falling proportion of the central dialect was significantly longer than its northern counterpart. The three-way interaction of DIALECT \times GENRE \times POSITION was also significant. This was due to the fact that for the final position of unscripted speech, there was little dialectal difference. For females, none of the effects were significant. Speakers of both dialects showed similar fall-to-rise ratios across genres and syllable positions.

Table 15. Summary of GAMM fitted to female f0 data using the first position in scripted speech from the northern dialect as the reference. Boldface indicates dialectal differences for tonal height (parametric coefficients) and contour shape (smooth terms). ‘**’ $p < .01$, ‘***’ $p < .001$.

Parametric coefficients	Estimate	S.E.	t value
(Intercept)	212.62	13.21	16.09***
GENRE1:POSITION2	-8.00	2.21	-3.62***
GENRE2:POSITION1	-1.24	9.85	-0.13
GENRE2:POSITION2	-15.23	10.06	-1.51
DIALECT2:GENRE1:POSITION1	-7.66	7.24	-1.06
DIALECT2:GENRE1:POSITION2	-9.66	7.23	-1.34
DIALECT2:GENRE2:POSITION1	-7.17	7.63	-0.94
DIALECT2:GENRE2:POSITION2	2.40	7.62	0.32
Smooth terms	edf	Ref.df	F
GENRE1:POSITION1	5.03	6.16	22.48***
GENRE1:POSITION2	5.45	6.63	67.00***
GENRE2:POSITION1	5.32	6.48	23.88***
GENRE2:POSITION2	5.25	6.40	37.87***
DIALECT2:GENRE1:POSITION1	1.00	1.00	0.00
DIALECT2:GENRE1:POSITION2	1.00	1.00	16.87***
DIALECT2:GENRE2:POSITION1	1.83	2.29	0.85
DIALECT2:GENRE2:POSITION2	1.80	2.24	0.26
SUBJECT	20.95	22.00	884.422***
ITEM	21.40	22.00	248.88

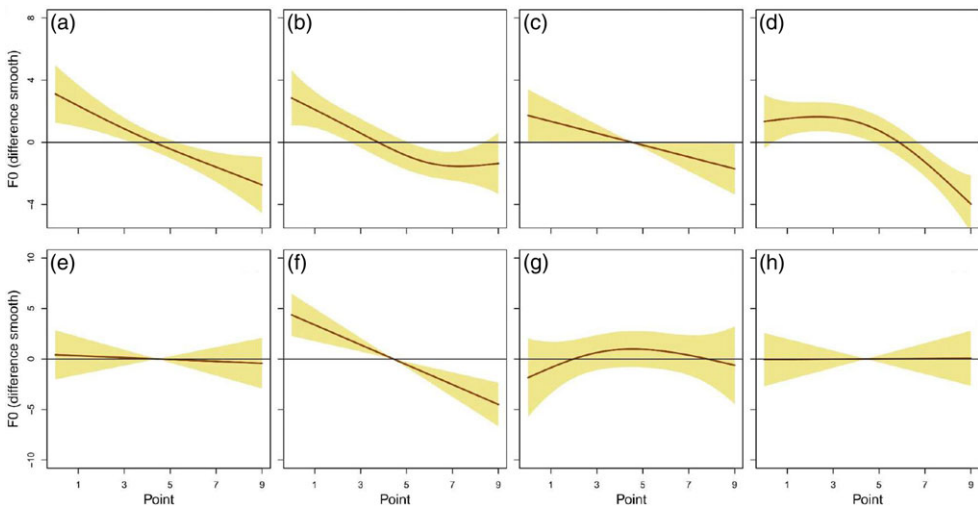


Figure 12. Difference smooths predicted by GAMM for males (a–d) and females (e–h) in four GENRE/POSITION treatment conditions using the northern variety as the reference. The first and second columns respectively represent the first and second syllables of the scripted speech, while the third and fourth columns respectively represent the first and second syllables of the unscripted speech. Colored areas indicate confidence intervals bounded by two standard errors. Dialectal difference is not significant when the confidence intervals contain 0.

Table 16. Fixed effects of the linear mixed model for the falling proportion for males using the nonfinal position in northerner scripted speech as the reference. ‘*’ $p < .05$.

	Estimate	S.E.	t value
(Intercept)	33.58	3.19	10.53***
DIALECT2	10.78	4.35	2.48*
GENRE2	4.23	4.43	0.96
POSITION2	2.24	2.94	0.76
DIALECT2:GENRE2	-1.44	6.00	-0.24
DIALECT2:POSITION2	3.91	3.81	1.03
GENRE2:POSITION2	3.93	3.96	0.99
DIALECT2:GENRE2:POSITION2	-12.65	5.29	-2.39*

4 Discussion

Results in this study were original and surprising, as some of the effects did not go as expected. In this section, we focus on two main findings. We first look at how the realization of Tone 1 was affected by the various factors examined and compared the patterns with what was predicted in the previous studies (Fu 1999; Wu 2003). Next, we turned to the dialectal variations in the realization of Tone 1 and discussed how they might constitute the impression of the general public on the central dialect.

4.1 Tone 1 realization

Previous studies regarding the central dialect showed that tonal realization is dependent on a variety of factors, including dialect, syllable position, gender, and genre (Wu 2003; Fu 1999). However, results in this study seemed to go blatantly against many of the previous predictions. In the following, we examined the effects of these factors accordingly.

4.1.1 Dialect effect

Although rising realizations were indeed found to be fairly common for Tone 1 in the central dialect, as was predicted from studies on other tones (Fu 1999; Wu 2003), it is not exclusively so. Instead, both the central and the northern varieties in this study showed a similar preference for a perceptually rising realization. The central variety did not realize more rises than its northern counterpart, either. If anything, the latter seemed to have a slightly higher (albeit nonsignificant) tendency of realizing Tone 1 as a rising tone than the former (72% vs. 66%), as is shown in Table 3. This implies that the general impression of the dialectal split could not be easily explained through straightforward differences in rising rates, and more intricate cues might be at work. The features of the central variety found in the previous literature thus require further examinations to allow direct comparisons between the two dialects so as to see whether qualitative differences do in fact exist.

4.1.2 Position effect

Rising contours were not exclusive to the utterance-final position, which is also contrary to previous findings and assumptions (cf. Fu 1999, Khoo 2020, and Wu 2003). Both the final and pre-final target CVN syllables elicited predominantly rising realizations, as shown in

Table 3. However, syllable position was still a key factor, as the rising realization decreased from 85% in the final position to 53% in the pre-final, which was more than a 30% decrease. The degree of the rise was also position-dependent (Figure 9). At the final position, male and female speakers showed an average excursion of 14.5 Hz and 26.1 Hz, respectively, but at the pre-final position, the degree of the rise was reduced to 7.45 Hz and 13.8 Hz, respectively, resulting in a decrease of more than 50%. In other words, rising Tone 1 is not only more common in the final position, but its rise is also more prominent.

Although not originally planned, syllable structure also seemed to play a position-dependent role in tonal realization. In the final position, the rising rate for non-CVN pairing syllables was 74%, which is fairly comparable to that of the CVN targets (Table 6). However, in the pre-final position, the rising rate plummeted to lower than 20%, and for males, it even dropped to downright zero, as none of their pre-final realizations were found to be rising (Figure 6). This implies that even though the rising contour was not exclusive to the final position, the location was still the most felicitous to a rising realization, and was relatively impervious to the effect of syllable structure.

We suspect that the final position was special because of its duration. Syllables in the final position are usually much longer than those in the pre-final positions due to final lengthening (Shen 1992; Lehiste 1975). Therefore, they can more easily accommodate rising realizations, which generally require longer syllable carriers than level ones (Ho 1976; Deng, Shi & Lu (2006), and also Figure 4). This could also explain why syllable structure played a larger role in the pre-final than the final position in rising realizations. A non-CVN pre-final syllable (CV in this case) had an average duration of 162 ms for its rising realization, which was 20% shorter than its 205 ms-long CVN counterpart in the same position. On the other hand, the difference was mitigated in the final position due to final lengthening, as the duration of a non-CVN syllable (CGV in this case) became fairly comparable to that of a CVN (279 ms vs. 277 ms). In other words, the position-dependent syllable effect likely stemmed from differential syllabic elasticity between pre-final and final positions. In the final position, final lengthening licenses long duration for syllables of all kinds, making them equally felicitous to rising Tone 1s, and thus obliterating the effect of syllable structure. On the other hand, as final lengthening is lifted in the pre-final position, the effect of syllable structure surfaces, and rising Tone 1s are more inclined to be realized on phonologically (and thus phonetically) longer syllables like CVN than shorter ones like CV and CGV. Similar accounts could also be applied to the position effect in previous studies (Fu 1999; Wu 2003). Since the dipping Tone 3 and the level Tone 4 are intrinsically longer than the low-falling Tone 3 and the high-falling Tone 4, respectively (Zee 1978; Xu & Sun 2002), they would naturally find the longer final syllables to be more facilitative to their realizations. In other words, if lengthened duration is the key to these more complex variant realizations in the central dialect, then one would expect syllables like CGVN to be the most attractive to the variants in the pre-final position, as it is phonologically the longest syllable in Mandarin.

4.1.3 Gender effect

Another discrepancy between the current results and those of previous findings lay in the effect of gender, as rising realizations were more common among females than males in this study, contrary to what was found previously with the other tones (cf. Fu 1999 and Wu 2003). The effect seemed fairly robust, and was consistently observed in both stimulus syllables of CVN and pairing syllables of CV and CGV. For the former, there was an average of 13% increase (Figure 5), and for the latter, a 19% increase was observed (Figure 6). There are potentially two reasons for such a discrepancy. One possibility might be a general tone-dependent attitude. While the central variety seems to show a ubiquitous preference

for ending high across all tones, the northern standard dialect might have adopted this preference for its Tone 1 only. This is plausible, since perception studies have shown that both high-level and high-rising contours are considered as adequate realizations of Tone 1 by native listeners (Massaro, Cohen & Tseng 1985). In other words, rising Tone 1 might have taken advantage of the allotonic variation that has already existed in the tone way before the central variety has come around. Therefore, the adoption of the rising rendition has not yet reached the level of consciousness and little negative connotation has been attached (Labov 2001). On the other hand, the dipping Tone 3 variant might have at least partially coincided with the emphatic form of the tone (Kubler 1985a, 1985b; Fon & Chiang 1999; Tsao 2000; Fon, Chiang & Cheung 2004) or the slight dipping rendition of Tone 2 (see Footnote 7), while the high-level Tone 4 likely bears a strong resemblance to the canonical high-level Tone 1, both of which should be well above the level of consciousness. Since female speakers tend to avoid negative variants of which they are consciously aware, they might hold different attitudes towards the three tones, preferring the rising Tone 1 variant, but dispreferring the dipping Tone 3 and the high-level Tone 4 variants.

Another possibility might be the perceptual acuity of the judges. Since female speakers in this study demonstrated a higher pitch range, a larger pitch excursion (Figures 8 & 11), and longer syllable duration than males (270 ms vs. 231 ms), it might be easier for the judges to perceive a rise in female than male speech. Liu (2013) found that the just noticeable differences (JNDs) for a Mandarin rising contour range from 4-10 Hz. However, this is modulated by pitch height (Jongman et al. 2017), and a higher pitch register raises perceptual sensitivity and lowers the JNDs. Detection of pitch excursion is also duration-dependent. Listeners generally need at least five complete cycles to detect pitch movements (Patterson, Peters & Milroy 1983), which means longer duration is required for correct identification of lower pitch excursions than higher ones (Fyk 1987). Since males generally spoke faster than females in this study, it is likely that some rising contours intended by the male speakers went undetected due to threshold limits. Two of the judges in the current study also mentioned that shorter tones were more difficult for them to hear the excursions.

4.1.4 Genre effect

The final discrepancy was the effect of genre. It was in general not very robust. For both dialects, unscripted speech did not elicit more rising contours than scripted speech (Table 3), which is surprising since it somewhat contradicts previous findings (cf. Fu 1999 and Wu 2003). One suspects that this has to do with the different materials used. Although both Fu (1999) and Wu (2003) included read and spontaneous speech as their stimuli, they used much longer utterances than those in the current study. For read speech, they used both isolated sentences and long passages, while for spontaneous speech, they used free conversations. In other words, their read and spontaneous speech differ not only in genre, but also in factors like syllable structure, utterance length, and focus placement, among others. In this study, we strove to control as much as possible for the potential extraneous factors and only varied genres. Since little effect was found, we could probably safely say that the effect of genre on the realization of rising Tone 1 either only surfaces as utterance length increases, or it really is not a robust factor at all. Previous claims for such an effect might thus be exclusive to other tonal categories, or due to factors other than genre that are not carefully controlled.

4.2 Dialectal variations

Although little dialectal difference was found in the rising rates, acoustic measurements did show some interesting results that were dialect-dependent, which might have contributed

to the percept of the dialectal divide. In addition to lower pitch register (Figure 8), which is consistent with previous findings (Fon et al. 2011; Khoo 2020), there are also some other cues that showed dialectal differences, including declination, upstep, and rising realization, which are detailed in the following sections.

4.2.1 Declination and upstep

The novel effect of dialect-dependent declination is rather interesting. For the northern variety, no trend of declination was observed. Regardless of genre, both the pre-final and the final syllables were of comparable pitch height (Figure 9). This was expected since the stimuli adopted in this study were merely short disyllabic words and declination is usually more observable in longer stretches of utterances (Shih 1988). What is surprising is the central dialect. Even for the short disyllabic stimuli, declination was still robust in scripted speech, and there was a significant downward trend of approximately 8 Hz for males and 11 Hz for females for the initial reference point. For unscripted speech, the trend was more complex and demonstrated a gender-dependent pattern. Male speakers still showed a similar downward trend while females demonstrated a reverse pattern. The effect of the decline was largely annihilated due to an upstep, as females raised their pitch floor in the final position to a height comparable to that in the pre-final position.

4.2.2 Rising realization

The phonetic realization of the rising Tone 1 also showed some dialectal variations. Although both varieties preferred slight dipping to pure rising contours, the central variety demonstrated a more extreme preference (Table 8), and this was true of both genders. Acoustically, the dipping rendition was also realized in a dialect- and gender-dependent fashion. Central males generally preferred an initial fall that was steeper and proportionately longer than that of their northern counterparts (Figures 11 & 12). The average slope for the central dialect was -99 Hz/s, as compared to the -80 Hz/s of the northern males.⁹ The falling arm occupied approximately 47% of the total duration for the central group, as compared to the 39% for the northern group. For females, the dialectal difference was smaller. Except for the second syllable of scripted speech, in which central females also showed a steeper initial fall than their northern counterparts (-123 Hz/s vs. -94 Hz/s), not much difference was found between the two dialects for other combinations of genre and position (central -173 Hz/s vs. northern -171 Hz/s). The falling proportions were also fairly comparable (central 41.4% vs. northern 40.7%). It is worth noting that even though the central dialect tended to show steeper initial falls than their northern counterparts, few have crossed the perceptual boundary between a rising and a dipping tone (Fon, Chiang & Cheung 2004), as the majority of the tokens were still regarded as rising realizations by the three judges, and none were deemed as dipping.

The central speakers' emphasis on the initial fall of the rising Tone 1 is especially intriguing. Although dipping realizations are fairly common in the realization of rising tones in Mandarin (Fon & Chiang 1999; Shi & Wang 2006; Fon 2020), the initial falling portion is usually considered a natural byproduct of physiological effort (Shi & Wang 2006), and is not of phonological significance (Chao 1956, 1968; Shih 1988). As a result, it is often left undiscussed in many acoustic studies (Ho 1976; Connell, Hogan & Rozsypal 1983; Shen 1990). However, this does not seem to have prevented listeners from making use of the section. Perception studies have shown that the initial fall serves as an effective cue to a rising tone

⁹ To accommodate for the nonlinearity of the falling portion and facilitate cross-dialectal comparisons, falling slopes were calculated from the section between Extraction Points 1 and 4 based on the 10-point pitch extraction contours in Figure 11, as most of the concave points fell beyond Point 4.

when the following rising portion is obliterated (Fon, Chiang & Cheung 2004). Although it is unclear whether listeners would actively utilize cues residing in the falling portion when the rising portion is readily available, they are apparently capable of doing so when situations arise. Therefore, the modification on the falling portion adopted by the central speakers could very likely have helped create the dialectal divide perceived by the general public.

4.2.3 Understanding the public impression

The peculiar declination pattern and realization of Tone 1 in the central dialect were unexpected and likely contributed to the impression of the general public. The steeper decline of the bottomline, along with lower pitch register, might have made the dialect sound distinctly lower than its northern counterpart. The extra emphasis on the initial falling portion of a rising Tone 1, manifested through a stronger preference for the dipping rendition and a steeper initial fall in both genders, and a proportionately longer initial fall among males, might have also helped. In other words, unlike what was suggested in previous studies (cf. Huang & Fon 2011 and Khoo 2020), the percept of having a lower pitch register was not achieved by merely a single cue, but was likely established through a combination of cues. By adopting more falls, steeper falls, and proportionately longer falls, the central dialect could effectively reach a much lower pitch level than its northern variety, further strengthening the public impression of the dialect being perceptually lower in pitch.

On the other hand, the final upstep of the central female speakers, coupled with the rising contour and the pitch range expansion commonly found in this position, likely constituted the percept of the ending high in unscripted speech. By raising the bottomline, central females could thus make the final rise perceptually more ear-catching and create a deeper impression of the dialectal flair since higher pitch register raises the JND of a rise (Jongman et al. 2017). The annihilation of the usual dialect-specific downward trend might have also created a marked cue for listeners from outside the variety to take a special note. In other words, the tendency to end high that has been deeply ingrained in the mind of the general public might not owe only to the rise itself, as the high-rising contour might be well within the boundaries of a typical Tone 1 when perceived in vacuum, and is perceived as anything but marked (Massaro, Cohen & Tseng 1985). Instead, the percept is likely created through an exact combination of an enlarged final rise in the foreground against a disruption of the downward trend in the background, as the latter provides an enhancing contrast for the former. Previous studies on the perception of coarticulated tones also found that accurate tonal identification relies not only on the contour of the tone in question, but also on the excursions of the preceding and following tones (Xu 1994; Qin & Zhang 2022). Tonal perception out of context usually arrives at a tonal category that is starkly different from that perceived within context. In other words, it is the precise combination of the two that allows listeners to perform speech recognition smoothly. One suspects that the dialectal characteristics of tones could also be realized in a similar fashion. Both the rising contour and its peculiar declination pattern might be essential in creating the general impression of the dialect.

4.2.4 Gender differences

The cues that could potentially contribute to the dialectal divide also showed a rather interesting gender-dependent pattern. As indicated in Table 17, all of the acoustic cues that central males used to demonstrate dialectal differences are related to creating an impression of a lower pitch register. Most of them were implemented forcefully disregarding any potential effects of genre and syllable position. On the other hand, females have adopted

Table 17. A summary of the cues utilized by central speakers that potentially contribute to their dialectal signature. Both a full (✓) and a half check mark (✓) indicate statistical significance between the two dialects. The former is used for effects that are independent of genre and syllable position, while the latter is for effects that are dependent on the two factors.

	Male	Female
(a) Potential cues to lower pitch register		
Lower mean pitch	✓	✓
Steeper declination	✓	✓
More dipping renditions	✓	✓
Steeper initial falls	✓	✓
Longer initial falls	✓	
(b) Potential cues to ending high		
Disruption of declination		✓

an additional cue that is potentially relevant to creating a perceptually prominent trend of ending high. The distribution of their cues was also more sensitive to genre and syllable position, and did not show a general pattern across all conditions. One possibility for this gender split might be differential connotations. Khoo (2020) claimed that pitch range narrowing is due to negative Min transfer, and Huang & Fon (2011) also found a correlation between pitch lowering and Min proficiency. As Min is currently a nonofficial language in Taiwan, with its usage mainly limited to interactions among friends and family members (Huang 1993), it is likely that the adoption of a lower pitch register for Tone 1 is tinted with a negative connotation. On the other hand, to the best of our knowledge, no link has been made between Min transfer and the tendency of ending high. Its adoption might have been a pure innovation by central speakers, which has gradually spread to the north. Although its origin is from a nonstandard dialect, perception studies from the 80s, approximately twenty years before the term *Taichung Qiang* “Taichung accent” became stable (Khoo 2020), showed that Mandarin listeners tend to categorize both high-level and high-rising contours as legitimate realizations of Tone 1 (Massaro, Cohen & Tseng 1985). Therefore, the newly arrived rising rendition might have gone under the phonological radar and have not yet acquired much negative connotation during the adoption process. Since female speakers are more inclined to acquire changes from below and less inclined to acquire stigmatized variants, as was claimed in Labov’s (2001) gender paradox, it is likely for them to become more advanced in the tendency of ending high, but less so in the low pitch register.

5 Conclusion

This study attempted to verify a public impression that has been around for some time regarding *Taichung Qiang* “Taichung accent”, i.e., whether there is indeed a general tendency for ending high in the central variety (Khoo 2020). Tone 1 was chosen as the target of study so as to complement previous findings regarding Tone 3 (Fu 1999) and Tone 4 (Wu 2003).

Results are both unsurprising and surprising. Unsurprisingly, we did find rising contours to be the predominant realization of Tone 1 for the central variety. Although Tone 1 is phonologically a high-level tone, more than 80% of the final Tone 1s were realized with

a rise. On the surface, this seemed to be the answer that we have been looking for, as it outright supported the public impression of the dialect. However, as one pried further, it became surprisingly clear that the cues that constitute the public conception could not be as straightforward as one had originally assumed, since rising realizations were also the most dominant for the northern variety and the nonfinal position, which have not been heard as ending high before. Therefore, more nuanced cues should have been at work instead.

Careful comparisons between the two dialectal regions suggested that the major differences likely lay in the pitch range and the acoustic realizations of the rising contours. The central variety generally occupied a lower pitch register along a steeper declination slope, creating a perceptual backdrop of lower pitch, and their female speakers annihilated the downward trend through an upstep in pitch register and adopted a much larger rise in the final position of unscripted speech, producing a contrast that was more noticeable than that of their northern and male counterparts. In other words, the central dialectal flair for Tone 1 likely does not lie in the rising contour of the tone alone. Instead, the public impression of the dialect is formed through a precise combination of the foreground rising and disruption of the background lowering. This cue package mainly appears in the final position of unscripted speech among females, confirming the special status of the utterance-final position found in previous studies (Fu 1999; Wu 2003).

The female lead in rising Tone 1 was unexpected, and it implies that the Tone 1 variant, although nonstandard in its origin, does not carry a negative connotation like their Tone 3 and Tone 4 counterparts. This might be due to the fact that the cue package adopted by the central variety is partially based on a neutral high-rising allotone that has been around for quite some time (Massaro, Cohen & Tseng 1985). However, since only production data was included in the current study, future perception experiments would be needed in order to verify whether the cue combination proposed here is indeed utilized by listeners in pinpointing the central dialect, and if so, how they are weighted against each other. On a broader scale, the potential interplay between pitch register and pitch contour could also enlighten us with regard to how exactly our perceptual mechanism works in tonal perception. A rising contour realized at a register higher than expected might be perceptually more marked than one that is realized at an expected lower register, which is likely the underlying cause for the general impression of the public regarding the central dialect.

Aside from necessary perception experiments, a production study involving longer utterances would also be interesting. As declination becomes more inevitable when utterances are lengthened, larger dialectal differences due to disparate slopes of the decline might thus surface. It would then be interesting to see whether the patterns observed in this study still hold. A judgment task would also be useful in determining whether central males and females are judged differently by speakers of other dialects with regard to the “accentedness” of their speech, and whether the differential patterning between the two genders could be reflected in the perception of a native ear. All of these would merit further studies.

Acknowledgments This paper was supported by a research grant awarded to the first author by the National Science Council, Taiwan (NSC95-2411-H-002-046-). Parts of this paper were presented at the 20th International Congress of Phonetic Sciences in Prague, Czech Republic in 2023. Many thanks to the helpful comments provided by Dr. Oliver Niebuhr and the two anonymous reviewers. They gave this manuscript a completely new look. Thanks to Hui-lu Khoo, Hsin-Yi Lin, and Hsuan-fang Chen for the help with subject recruitment and recording, and Yen-chen Lu for the help with tonal judgment. Special thanks to all the subjects. Naturally, all the faults are ours.

Appendix

The following table is the full stimulus list used in the experiment.

	Target Position	
	First	Second
ban¹	ban¹jia¹ ‘to move (to a new residence)’ ban¹ji² ‘a class (in school)’ ban¹jiang³ ‘to award’ ban¹dai⁴ ‘class representative’	fen¹ban¹ ‘to group students into classes’ tong²ban¹ ‘in the same class’ wan³ban¹ ‘night shift’ que⁴ban¹ ‘freckle’
dan¹	dan¹xin¹ ‘to worry’ dan¹chun² ‘simple, naive’ dan¹dian³ ‘à la carte’ dan¹ji⁴ ‘unit price’	xian¹dan¹ ‘elixir’ ming²dan¹ ‘roster’ bang³dan¹ ‘admission list’ fu⁴dan¹ ‘burden’
gan¹	gan¹die¹ ‘godfather’ gan¹ke² ‘dry cough’ gan¹xi³ ‘dry cleaning’ gan¹ga⁴ ‘embarrassed, awkward’	zhu¹gan¹ ‘pork liver’ lan²gan¹ ‘railing’ bing³gan¹ ‘cookie’ dou⁴gan¹ ‘dried tofu’

References

- Alexander, Jennifer Alexandra. 2010. *The theory of adaptive dispersion and acoustic-phonetic properties of cross-language lexical-tone systems*. Ph.D. dissertation, Northwestern University.
- Barr, Dale J., Roger Levy, Christoph Scheepers & Harry J. Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68(3), 255–278.
- Bates, Douglas, Martin Mächler, Ben Bolker & Steven Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1), 1–48.
- Boersma, Paul & David Weenink. 2009. *Praat: Doing phonetics by computer* (version 5.1). <http://www.praat.org/>.
- Chao, Yuan Ren. 1956. Tone, intonation, singsong, chanting, recitative, tonal composition, and atonal composition in Chinese. In Morris Halle (ed.), *For Roman Jakobson*, 52–59. The Hague: Mouton.
- Chao, Yuan Ren. 1968. *A Grammar of Spoken Chinese*. Berkeley: University of California Press.
- Chen, Sheng H. 2005. The effects of tones on speaking frequency and intensity ranges in Mandarin and Min dialects. *The Journal of the Acoustical Society of America* 117, 3225–3230.
- Chuang, Yu-Ying, Janice Fon, Ioannis Papakyritsis & Harald Baayen. 2021. Analyzing phonetic data with generalized additive mixed models. In Martin J. Ball (ed.), *Manual of Clinical Phonetics*, 108–138. London & New York: Routledge.
- Chung, Karen Steffen. 2006. Hypercorrection in Taiwan Mandarin. *Journal of Asian Pacific Communication* 16(2), 197–214.
- Connell, Bruce A., John T. Hogan & Anton J. Rozsypal. 1983. Experimental evidence of interaction between tone and intonation in Mandarin Chinese. *Journal of Phonetics* 11(4), 337–351.
- Deng, Dan, Feng Shi & Shinan Lu. 2006. The contrast on tone between Putonghua and Taiwan Mandarin. *Acta Acustica* 31(6), 536–541.
- Fon, Janice. 2020. The phonetic realizations of the Mandarin phoneme inventory: The canonical and the variants. In Huei-mei Liu, Feng-ming Tsao & Ping Li (eds.), *Speech Perception, Production and Acquisition: Multidisciplinary Approaches in Chinese Languages*, 11–36. Singapore: Springer.
- Fon, Janice & Wen-Yu Chiang. 1999. What does Chao have to say about tones? – A case study of Taiwan Mandarin. *Journal of Chinese Linguistics* 27(1), 15–37.
- Fon, Janice, Wen-Yu Chiang & Hintat Cheung. 2004. Production and perception of two dipping tones (T2 and T3) in Taiwan Mandarin. *Journal of Chinese Linguistics* 32(2), 249–280.
- Fon, Janice, Jui-Mei Hung, Yi-Hsuan Huang & Hui-Ju Hsu. 2011. Dialectal variations on syllable-final nasal mergers in Taiwan Mandarin. *Language and Linguistics* 12(2), 273–311.

- Fu, Jo-Wei. 1999. *Chinese tonal variation and social network – A case study in Tantz Junior High School Taichung, Taiwan*. M.A. thesis, Providence University.
- Fyk, Janina. 1987. Duration of tones required for satisfactory precision of pitch matching. *Bulletin of the Council for Research in Music Education* 91, 38–44.
- Ho, Aichen Ting. 1976. The acoustic variation of Mandarin tones. *Phonetica: International Journal of Speech Science* 22, 353–367.
- Huang, Shuanfan. 1993. *Language, Society, and Ethnic Identity: A Study on Language Sociology in Taiwan*. Taipei: Crane.
- Huang, Yi-Hsuan & Janice Fon. 2011. Investigating the effect of Min on dialectal variations of Mandarin tonal realization. In Wai-Sum Lee & Eric Zee (eds.), *Proceedings of the 17th International Congress of Phonetic Sciences (ICPhS XVII)*, City University of Hong Kong, 918–921.
- Johnson, Keith. 2004. Massive reduction in conversational American English. In Kiyoko Yoneyama & Kikuo Maekawa (eds.), *Spontaneous Speech: Data and Analysis. Proceedings of the 1st Session of the 10th International Symposium*, 29–54. The National International Institute for Japanese Language.
- Jongman, Allard, Zhen Qin, Jie Zhang & Joan A. Sereno. 2017. Just noticeable differences for pitch direction, height, and slope for Mandarin and English listeners. *The Journal of the Acoustical Society of America* 142(2), EL163–169.
- Khoo, Hui-lu. 2020. A preliminary study of the tonal features of central Taiwan Mandarin. *Taiwan Journal of Linguistics* 18(1), 115–157.
- Kösling, Kristina, Gero Kunter, Harald Baayen & Ingo Plag. 2013. Prominence in triconstituent compounds: Pitch contours and linguistic theory. *Language and Speech* 56(Pt 4), 529–554.
- Kubler, Cornelius C. 1985a. *The Development of Mandarin in Taiwan: A Case Study of Language Contact*. Taipei: Student Book.
- Kubler, Cornelius C. 1985b. The influence of Southern Min on the Mandarin of Taiwan. *Anthropological Linguistics* 27, 156–176.
- Labov, William. 1972. *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press.
- Labov, William. 2001. *Principles of Linguistic Change, Vol. 2: Social Factors*. Oxford: Blackwell.
- Lehiste, Ilse. 1975. The role of temporal factors in the establishment of linguistic units and boundaries. In Wolfgang. U. Dressler & F. V. Mareš (eds.), *Phonologica 1972: Akten der zweiten internationalen Phonologie-Tagung*, 115–122. München-Salzburg, Germany: Wilhelm Fink Verlag.
- Liljencrants, Johan & Björn Lindblom. 1971. Numerical simulation of vowel quality systems: The role of perceptual contrast. *Language* 48(4), 839–862.
- Liu, Chang. 2013. Just noticeable difference of tone pitch contour change for English- and Chinese-native listeners. *The Journal of the Acoustical Society of America* 134(4), 3011–3020.
- Massaro, Dominic W., Michael M. Cohen & Chiu-yu Tseng. 1985. The evaluation and integration of pitch height and pitch contour in lexical tone perception in Mandarin Chinese. *Journal of Chinese Linguistics* 13(2), 267–289.
- National Statistics R.O.C. 2021. *109 Nian Renkou Ji Zhuzhai Pucha Zongbaogao Tiyaoy Fenxi [The 2020 Population and Housing Census: A general summary report on the statistic results and analyses]*. Taipei: Ministry of Interior R.O.C. Taiwan. <https://www.stat.gov.tw/News.aspx?n=2750&sms=11062>.
- Patterson, Roy D., Robert W. Peters & Robert Milroy. 1983. Threshold duration for melodic pitch. In Rainer Klinke & Rainer Hartmann (eds.), *Hearing – Physiological Bases and Psychophysics*, 321–326. Berlin: Springer.
- Qin, Zhen & Jingwei Zhang. 2022. The use of tonal coarticulation cues in Cantonese spoken word recognition. *JASA Express Letters* 2(3), 035202.
- R Core Team. 2021. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
- Shen, Xiaonan Susan. 1990. Tonal coarticulation in Mandarin. *Journal of Phonetics* 18(2), 281–295.
- Shen, Xiaonan Susan. 1992. A pilot study on the relation between the temporal and syntactic structures in Mandarin. *Journal of the International Phonetic Association* 22 (1/2), 34–43.
- Shi, Feng & Ping Wang. 2006. A statistic analysis of the tones in Beijing Mandarin. *Studies of the Chinese Language*, 310, 33–40.
- Shih, Chi-Lin. 1988. Tone and intonation in Mandarin. *Working Papers of the Cornell Phonetics Laboratory* 3, 83–109.
- Shih, Michael. 2011. *Huandao Lyuxing [Traveling around the Island]*. Ho Vision Entertainment.
- Sun, Yan & Chilin Shih. 2021. Boundary-conditioned anticipatory tonal coarticulation in standard Mandarin. *Journal of Phonetics* 84, 101018.
- Tomaschek, Fabian, Benjamin V. Tucker, Matteo Fasiolo & R. Harald Baayen. 2018. Practice makes perfect: The consequences of lexical proficiency for articulation. *Linguistics Vanguard* 4(s2), 20170018.
- Trudgill, Peter. 1972. Sex, covert prestige, and linguistic change in the urban British English of Norwich. *Language in Society* 1, 179–95.
- Tsao, Feng-Fu. 2000. Taiwanized Japanese and Taiwan Mandarin – Two case studies of language contact during the past hundred years in Taiwan. *Hanxue Yanjiu (Chinese Studies)* 18(1), 273–297.

- Wieling, Martijn, Fabian Tomaschek, Denis Arnold, Mark Tiede, Franziska Bröker, Samuel Thiele, Simon N. Wood & R. Harald Baayen. 2016. Investigating dialectal differences using articulatory data. *Journal of Phonetics* 59, 122–143.
- Wood, Simon N. 2011. Fast stable restricted maximum likelihood and marginal likelihood estimation of semi-parametric generalized linear models. *Journal of the Royal Statistical Society. Series B: Statistical Methodology* 73(1), 3–36.
- Wood, Simon N. 2017. *Generalized Additive Models: An Introduction with R*. 2nd ed. New York: Chapman and Hall/CRC Press.
- Wu, E-Chin. 2009. *The effect of Min proficiency on the realization of tones and foci in Taiwan Mandarin-Min bilinguals*. M.A. thesis, National Taiwan University.
- Wu, Shu-Juan. 2003. *A sociolinguistic study of Chinese tonal variation in Puli, Nantou Taiwan*. M.A. thesis, Providence University.
- Xu, Yi. 1994. Production and perception of coarticulated tones. *The Journal of the Acoustical Society of America* 95(4), 2240–2253.
- Xu, Yi. 1997. Contextual tonal variations in Mandarin. *Journal of Phonetics* 25(1), 61–83.
- Xu, Yi & Xuejing Sun. 2002. Maximum speed of pitch change and how it may relate to speech. *The Journal of the Acoustical Society of America* 111(3), 1399–1413.
- Zee, Eric. 1978. Duration and intensity as correlates of F0. *Journal of Phonetics* 6(3), 213–220.

Cite this article: Fon Janice and Chuang Yu-Ying (2025). When a rise is not only a rise: An acoustic analysis of the impressionistic distinction between northern and central Taiwan Mandarin using Tone 1 as an example. *Journal of the International Phonetic Association*. <https://doi.org/10.1017/S0025100324000100>