

# Positional and Phonotactic Effects on the Realization of Taiwan Mandarin Tone 2

Janice Fon & Hui-ju Hsu

Department of English  
National Taiwan Normal University  
[jfon@cc.ntnu.edu.tw](mailto:jfon@cc.ntnu.edu.tw), [hsuhj.hsu@msa.hinet.net](mailto:hsuhj.hsu@msa.hinet.net)

## Abstract

This study investigates how phonotactics and sentential positioning affect the realization of Taiwan Mandarin Tone 2. 10 Taipei speakers read sentences containing target syllables of different syllable types in isolation and in various sentential positions. Results showed that comparing across five different phonotactic types, N-initial syllables were relatively high at the onset, the turning point, and the offset while G-initial syllables were low at all three points. V-initial syllables were high at the turning point and the offset while obstruent-initial syllables ended with relatively low pitch. In general, the offset is higher than the onset point when the syllable was read in isolation or sentence-finally. However, in sentence-initial and -medial positions, the onset tended to be higher in pitch than the offset, especially in female speech. V-initial syllables tended to have higher offsets than onsets regardless of positions. In terms of duration, the falling ratio is the highest sentence-medially. With L-initial and N-initial syllables, the falling ratio was over 50% of the total syllable sentence-medially and -finally. There was no difference in the degree of steepness between the falling and the rising portions of the tone, but female speakers in general showed steeper slopes than male, especially in sentence-final positions.

## 1. Introduction

Like its mainland counterpart (Putonghua), Taiwan Mandarin (Guoyu) has four tones, traditionally termed Tone 1 (T1), Tone 2 (T2), Tone 3 (T3), and Tone 4 (T4). Although in Putonghua, T2 is realized as mid-rising ([1] & [2]), in Guoyu, however, instead of the prescriptive rising contour, T2 has a dipping shape and occupies the mid pitch range, which is very much like T3 ([3], [4], & [5]) (Figure 1). However, a closer examination shows that there are still minor differences ([3] & [4]). In isolation, T2 occupies a higher pitch range, has a higher offset than onset, a longer rising portion than falling, and shallower slopes. On the other hand, a full T3 occupies a lower pitch range, has a higher onset than offset, a longer falling portion than rising, and steeper slopes.

In this study, we would like to examine how phonotactic and positional factors influence the realization of Guoyu T2. Previous studies showed that the dipping contour is less likely to appear in syllables with bare onsets [3]. Using trisyllabic words, Fon ([3] & [4]) also found a position effect. The subject proportionately lengthened the falling contour as the syllable is positioned towards the end of the word.

Since Fon ([3] & [4]) used only one subject for her study, it is unclear whether the realization of T2 and the related interaction concerning such are idiosyncratic or can be generalized to the overall Guoyu-speaking population.

Therefore, we would like to look into this issue in a more systematic manner so that possible generalization can be made.

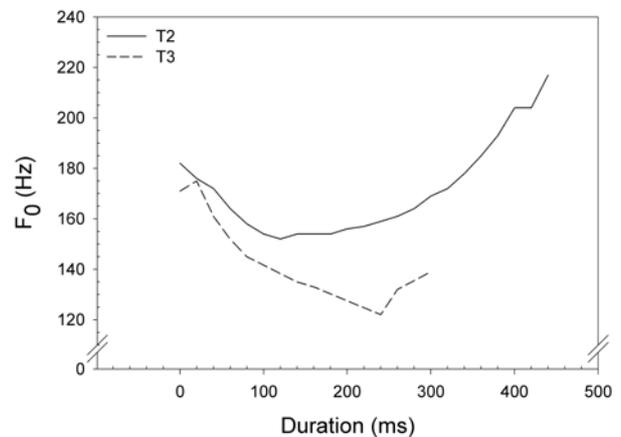


Figure 1: Guoyu T2 & T3.

## 2. Methods

### 2.1. Subjects

10 subjects, five male and five female, who were native speakers of Guoyu and were born and raised in the Taipei area were recruited (age:  $\bar{X} = 22.10$ ,  $SD = 1.76$ ). The subjects were paid with monetary rewards. There are two reasons for choosing Taipei Guoyu speakers. One is to make the current study comparable to Fon's previous studies ([3] & [4]), since her subject was a native Taipei speaker. The other is because Taipei Guoyu is considered the standard variety in Taiwan and whether and how T2 realization is influenced by phonotactic and positional factors has theoretical and pedagogical significance.

### 2.2. Stimuli

Five different phonotactic types were included in the study. They were obstruent-initial (C-initial, e.g., *pi2* 'skin'), glide-initial (G-initial, e.g., *ya2* 'tooth'), liquid-initial (L-initial, e.g., *liu2* 'flow'), nasal-initial (N-initial, e.g., *mian2* 'cotton'), and bare onset words (V-initial, e.g., *wu2* 'none'). There were six different syllables in each category. The syllables were written in traditional Chinese characters printed on index cards and were placed either in isolation or in a carrier sentence. Table 1 illustrates the carrier sentences of different sentential positions.

In total, there were 6 (tokens)  $\times$  5 (syllable types)  $\times$  4 (positions) = 120 T2 stimuli.

Table 1: Carrier sentences of different sentential positions

Position	Carrier Sentence
Isolation	<i>X</i> 'X'
S-initial	<i>X</i> zhege zi hen nan nian. 'X' word is hard to pronounce.
S-medial	Zhege <i>X</i> zi hen nan nian This 'X' word is hard to pronounce.
S-final	Zhege zi shi nian <i>X</i> . This word is pronounced as 'X'.

### 2.3. Equipment

Recordings were done using a SONY PCM-M1 Digital Audio Recorder with Maxell R-64DA 60 min DAT tapes and a SHURE SM10A head-mounted microphone.

### 2.4. Procedure

The recordings took place in the Phonetics Laboratory of the NTNU. Subjects were shown the index cards one by one and were asked to read out loud the isolated syllables and sentences on the cards using natural intonation. The cards were ordered so that they were semi-randomized for each subject and the index cards containing single syllables were read before those containing sentences. Subjects were allowed to rest in-between if they needed to. The whole process took about 15 min. The recording was done at a sampling rate of 48 kHz and a D-to-D transfer was then done at a downsampled sampling rate of 22050 Hz for further analyses.

### 2.5. Measurements

A Praat script was written for automatic pitch extraction and pitch tracks were later hand-checked and -corrected for doubling and halving. Since most of the syllables collected were dipping (only 4 out of 1200 were non-dipping), three measurement points were extracted using a Praat script—the initial high (onset), the medial low (turning point), and the final high (offset) (Figure 2). The duration between adjacent reference points were also measured and the slope between them calculated (Hz/sec).

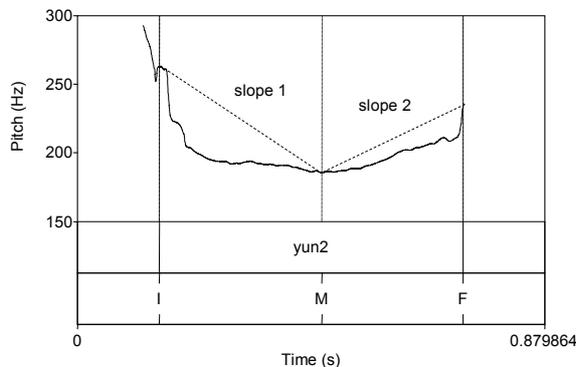


Figure 2: An example of pitch point extraction.

## 3. Results

Three sets of analyses were done on the data, including pitch, duration, and slope.

### 3.1. Pitch

A Position (4)  $\times$  Syllable type (5)  $\times$  Reference point (3)  $\times$  Gender (2) 4-way mixed ANOVA was done on  $f_0$  values. Results showed that all of the main effects were significant [P:  $F(1.96, 105.55) = 230.67, p < .0001, \eta^2 = .81$ ; S:  $F(2.97, 160.28) = 3.57, p < .05, \eta^2 = .06$ ; R:  $F(1.71, 88.45) = 389.35, p < .0001, \eta^2 = .88$ ; G:  $F(1, 54) = 788.67, p < .0001, \eta^2 = .94$ ].<sup>1</sup> Four of the 2-way interactions and two of the three-way interactions were also significant [P  $\times$  G:  $F(3, 162) = 43.65, p < .0001, \eta^2 = .45$ ; P  $\times$  R:  $F(2.67, 143.92) = 39.64, p < .0001, \eta^2 = .42$ ; S  $\times$  R:  $F(6.28, 339.07) = 6.14, p < .0001, \eta^2 = .10$ ; R  $\times$  G:  $F(2, 108) = 80.86, p < .0001, \eta^2 = .60$ ; P  $\times$  R  $\times$  G:  $F(6, 327) = 17.11, p < .0001, \eta^2 = .24$ ; P  $\times$  S  $\times$  R:  $F(14.11, 761.92) = 3.49, p < .0001, \eta^2 = .06$ ]. The four-way interaction was also significant [ $F(24, 1296) = 2.75, p < .0001, \eta^2 = .05$ ].

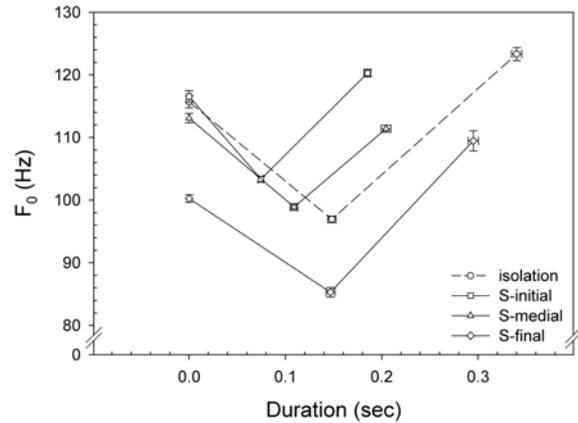


Figure 3: The P  $\times$  R effect in male speakers.

Since there were significant interaction effects, separate post-hoc analyses of Position  $\times$  Syllable type  $\times$  Reference Point regarding gender were done. For male speakers, all of the main effects were significant [P:  $F(2.41, 62.65) = 87.52, p < .0001, \eta^2 = .71$ ; S:  $F(4, 104) = 4.78, p < .01, \eta^2 = .16$ ; R:  $F(2, 52) = 208.44, p < .0001, \eta^2 = .89$ ]. Two of the 2-way interactions were also significant [P  $\times$  R:  $F(2.90, 75.38) = 17.10, p < .0001, \eta^2 = .40$ ; S  $\times$  R:  $F(8, 208) = 6.32, p < .0001, \eta^2 = .20$ ].

Post-hoc analyses using Bonferroni's adjustments showed that when T2 was placed in isolation or in sentence-initial or -final positions, the offset was always significantly higher than the onset ( $p < .001$ ) (Figure 3). However, in sentence-medial positions, the onset was higher than the offset, although this was non-significant. Phonotactics also made a difference (Figure 4). N-initial syllables tended to have high  $f_0$  for the

<sup>1</sup> Throughout the paper,  $df$ s were corrected using Huynh-Feldt's adjustments when Mauchly's test of sphericity was significant.

three reference points ( $p < .05$ ) while G-initial syllables tended to have low for all three ( $p < .01$ ). V-initial syllables tended to have low onsets, high turning points, and high offsets ( $p < .01$ ) and C-initial syllables tended to have the lowest offsets ( $p < .0001$ ).

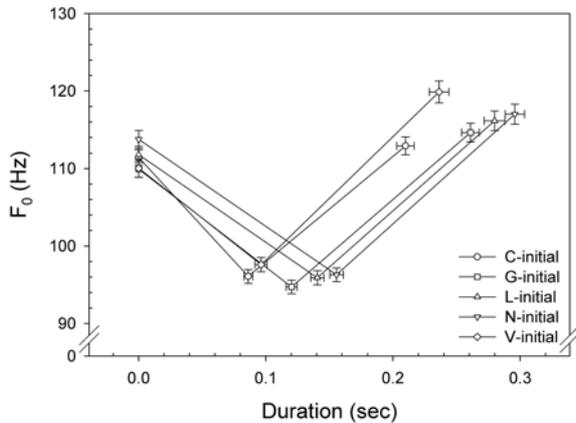


Figure 4: The S × R effect in male speakers.

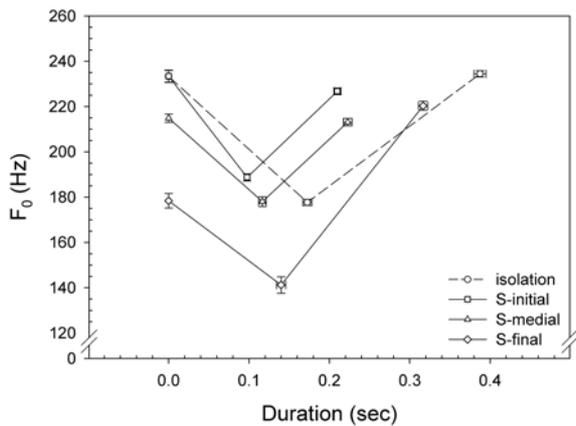


Figure 5: The P × R effect in female speakers.

The post-hoc ANOVA analysis on female speakers showed that two of the main effects were significant [P:  $F(1.68, 47.07) = 158.84, p < .0001, \eta^2 = .85$ ; R:  $F(1.67, 46.68) = 254.59, p < .0001, \eta^2 = .90$ ]. Two of the two-way interactions and one of the three-way interaction were also significant [P × R:  $F(2.54, 71.08) = 32.01, p < .0001, \eta^2 = .53$ ; S × R:  $F(6.42, 179.64) = 3.82, p < .001, \eta^2 = .12$ ; P × S × R:  $F(14.92, 417.71) = 3.67, p < .0001, \eta^2 = .12$ ].

Post-hoc analyses using Bonferroni's adjustments showed that offsets tended to be higher than onsets for syllables in isolation and in sentence-final positions ( $p < .0001$ ) (Figure 5). For sentence-initial and -medial syllables, offsets were lower than onsets ( $p < .05$ ). However, V-initial words tended to have higher offsets than onsets regardless of sentential positions (Figure 6).

In terms of phonotactic effects, V-initial syllables had the highest turning points and offsets while G-initial and C-initial syllables tended to have low values at these two points ( $p < .05$ ) (Figure 7). There was no significant difference among the syllable types at the onset position.

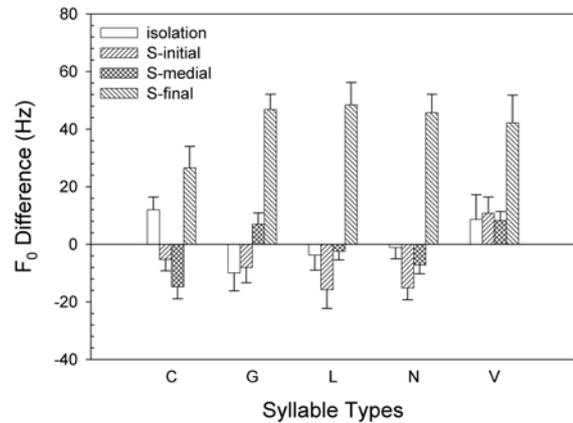


Figure 6: F<sub>0</sub> difference between onsets and offsets (offset – onset) in female speakers.

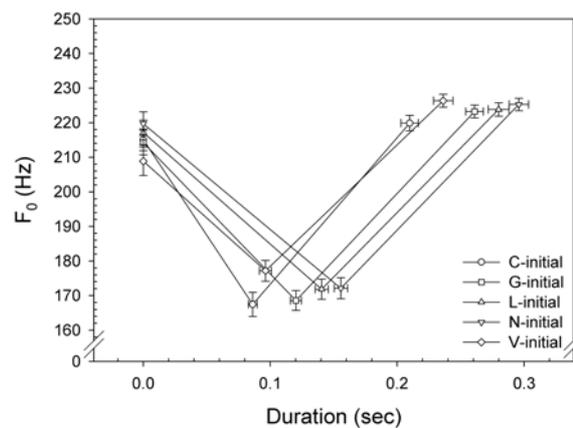


Figure 7: The S × R effect in female speakers.

### 3.2. Duration

A Position (3) × Syllable type (5) × Gender (2) 3-way mixed ANOVA was done on the arcsine-transformed falling ratios. Since Gender did not appear as a significant factor, a 2-way Position (3) × Syllable type (5) within-subject ANOVA was done instead to increase power. Results showed that all of the effects were significant [P:  $F(2.28, 134.49) = 19.39, p < .0001, \eta^2 = .25$ ; S:  $F(4, 236) = 33.17, p < .0001, \eta^2 = .36$ ; P × S:  $F(10.67, 629.75) = 2.62, p < .01, \eta^2 = .04$ ].

Post-hoc analyses using Bonferroni's adjustments showed that in general, N-initial and L-initial syllables have the longest falling portion while C-initial and V-initial syllables have the shortest ( $p < .01$ ). Syllables in sentential-medial positions

tended to have longer falling portions as compared to those in other positions ( $p < .05$ ) (Figure 8). This is especially true for N-initial and L-initial syllables, the falling portion of which even exceeded 50% of the total duration in sentence-medial and -final positions.

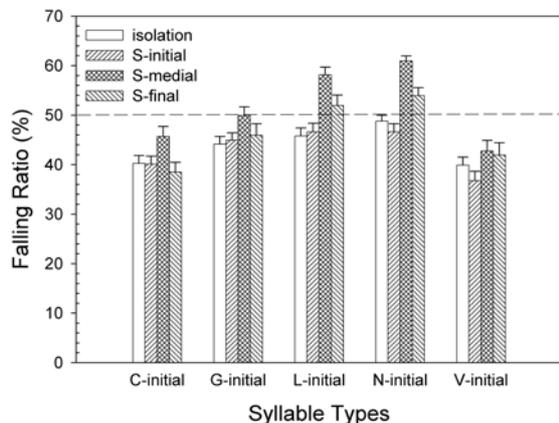


Figure 8: The  $P \times S$  interaction effect.

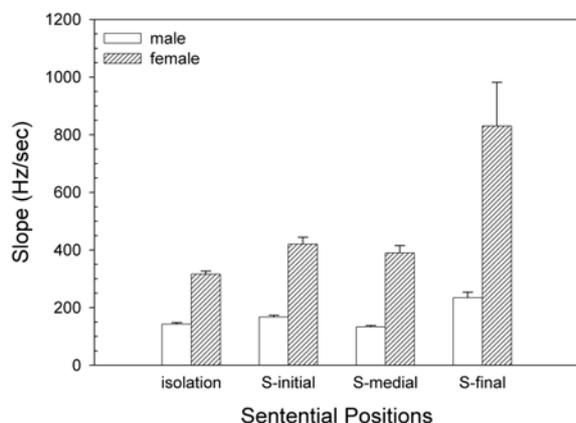


Figure 9: The  $P \times G$  interaction effect.

### 3.3. Slope

A Position (3)  $\times$  Syllable type (5)  $\times$  Direction (2: falling vs. rising)  $\times$  Gender (2) four-way mixed ANOVA was done on the absolute value of slope. Since Direction and Syllable type did not have any significant effect, groups were combined and a Position (3)  $\times$  Gender (2) 2-way mixed ANOVA was done instead. Results showed that all three effects were significant [P:  $F(1.02, 605.70) = 6.30, p < .05, \eta^2 = .01$ ; G:  $F(1, 594) = 18.81, p < .0001, \eta^2 = .03$ ;  $P \times G: F(3, 1782) = 3.41, p < .05, \eta^2 = .01$ ]. Post-hoc analyses using Bonferroni's adjustments showed that female speakers tended to have steeper slopes than male speakers ( $p < .0001$ ) (Figure 9). Syllables in isolation and in sentence-medial positions tended to have shallower slopes than sentence-initials and -finals ( $p < .05$ ).

## 4. Discussion

The findings in this study are interesting. First of all, dipping seems to have become the default contour for Guoyu T2, confirming our previous findings ([3] & [4]). Of the 1200 syllables collected, only 4 were nondipping. The shift of T2 from a rising to a dipping might be why the full form of T3 in Guoyu is somewhat different from that in Putonghua, the former being a dipping with a major fall and a minor rise, and the latter being a dipping with a major fall and a major rise, perhaps to better differentiate the two tones. Secondly, we also found that under certain conditions, T2 and T3 become more alike. Positionally, sentence-medial T2 syllables tend to have contours that are shaped more like T3 in that onsets were higher than offsets and falling portions were longer than rising portions. This is in contrast with the previous findings ([3]), as T2 in word-final positions were shown to be more like T3. The discrepancy between Fon ([3]) and the current study might be due to scope differences, as Fon examined only trisyllabic words in isolation, and the current study looks at monosyllables in various sentential positions. Phonotactically, this study corroborates previous findings in that V-initial syllables are not as pliable to this contour change. Of the five syllable types examined, V-initial words were the least similar to the canonical full T3 in Guoyu. In contrast, L-initial and N-initial syllables were more malleable, perhaps due to their phonetic properties, as both types of onsets can be lengthened to a significant extent.

## 5. Conclusions

This study confirmed the previous findings ([3] & [4]) that dipping contour seems to have become the default pattern for Taipei Guoyu T2. Furthermore, the contour of T2 changes with phonotactics and sentential positioning. Under certain environments, the distinction between T2 and T3 seems to be somewhat neutralized, hinting an ongoing leveling process of Guoyu tones.

## 6. Acknowledgements

This study could not have been possible if it were not for our diligent assistants, Sheau Yun Jamie Lin, Sally Hsuan-fang Chen, Sarah Yu-ying Chuang, and Yi-An Luo. Naturally, all the faults are ours.

## 7. References

- [1] Chao, Y.-R. Tone, intonation, singsong, chanting, recitative, tonal composition, and atonal composition in Chinese. In M. Halle (ed.) *For Roman Jakobson*, 52-59. Mouton: The Hague, 1956.
- [2] Chao, Y.-R. *A Grammar of Spoken Chinese*. University of California: Berkeley & Los Angeles, 1968.
- [3] Fon, Y.-J. J. *What are Tones Really Like?---An Acoustic-based Study of Taiwan Mandarin Tones*. Thesis. National Taiwan University, 1997.
- [4] Fon, J. and Chiang, W.-Y. What does Chao have to say about tones? ---a case study of Taiwan Mandarin. *Journal of Chinese Linguistics*, 27(1): 15-37, 1999.
- [5] Shen, X. S. and Lin, M. A perceptual study of Mandarin Tone 2 and 3. *Language and Speech*, 34(2): 145-156, 1991.