Speech Prosody 2024
2-5 July 2024, Leiden, The Netherlands

Question Intonation in Guanzhong Mandarin

Jiarui Zhang[1]

[1]University of Oxford
jiarui.zhang@ling-phil.ox.ac.uk

Abstract

This study examines the intonational tunes of syntactically unmarked polar questions in Guanzhong Mandarin (GuanM), particularly focusing on the exploration of the boundary tone in Chinese languages and the interaction between tone sandhi rules and intonation. We conducted a production study with eight subjects on disyllabic words. The results show that the question intonation in GuanM has a higher and raised register, compared to the statement intonation, given higher F0 mean and higher F0 maximum and minimum in questions. In addition, the longer duration of the last syllable of questions but not the first syllable correlates with the established definition of boundary tones, which typically occur on the last syllable of the intonational phrase. Finally, in the T2T2 question intonation, the F0 change rate in the second syllable was higher, while in T1T1, it was lower, compared to statement counterparts, highlighting the necessity and a characteristic of the high boundary tone (H%) in the question intonation of Chinese dialects that it moderates the extent of the falling tone and facilitates the rising tone rising.

Index Terms: Tone-Tune interaction, question intonation, phonology in Chinese dialects

1. Introduction

Guanzhong Mandarin (GuanM), a northern sub-dialect within the Mandarin group of Chinese dialects, is predominantly spoken in Shaanxi Province, China. While GuanM shares most of the segmental and syntactic features with Standard Mandarin, it differs in its tone contours and tone sandhi. In GuanM, there are four lexical tones: two falling tones- one with a low register (T1, 31) and one with a high register (T3, 53), a rising tone (T2, 24), and a high-level tone (T4, 55), along with a neutral tone. The tone sandhi rule in GuanM is listed as follows: T1 + T1 → T2 + T1: HL → LH (24) /__ HL, in which there is a disyllabic sequence of T1T1, the first falling T1 changes to a rising tone T2, while the second T1 remains unchanged, such as in words like [ʈʂhəŋ31→24 ɕjen31] "center", [kwei31→24 tɕia31] "nation", [faŋ31→24 fa31] "method", [tɕi31→24 liɛ31] "intense", and [kʰæ31→24 xua31] "blossom". Gong [1] argues that if the second tone is reduced to a neutral tone, which is low tone, the first tone will not undergo such a change. Examples include [kwəŋ31 tɕi31→1] "rooster", and [tɕi31 tɕyɤ31→1] "corner", where the first syllable maintains its base tone value.

Various languages use different methods for marking questions. Syntactic features play a significant role, such as using question words in Polish [2] and Russian [3], changing word orders in German [4] and French [5] and adding question particles in Mandarin Chinese [6] and Japanese [7]. Another common approach involves prosodic means, in which intonation becomes crucial for distinguishing questions from statements, especially in yes-no questions without syntactic markers [8]. Bolinger [9] suggests that high-pitched questions are a universal feature in intonation across languages. However, Rialland [10] summarizes two types of yes-no questions marking from 78 African languages: high-pitched yes-no question markers (including cancellation or reduction of downdrift, register expansion, rising of last Hs, cancellation or reduction of final lowering, final high tone or rising tune and final HL melody) and non-high-pitched ones (including final low tone or falling tune, final polar tone or M tone, lengthening, breathy termination, cancellation of penultimate lengthening and open vowel).

As for intonational studies in Chinese dialects, in general, previous research (e.g., [11], [12]) largely supports the findings that the intonation-induced F0 mainly affects F0 height rather than F0 contours, such that in questions, the height of F0 contours are higher, while the basic shape of F0 contours is maintained. However, debates exist, particularly about the temporal effect of intonation. One theory is the global-rising theory, which believes of a global rising over question sentences than statement counterparts (e.g., [13], [12]). An alternative perspective is the final-rising theory, which argues that at the local level, the F0 rise in questions is more visible in the final positions of the sentences (e.g., [14]).

The primary focus of this study is to investigate the tune of syntactically unmarked yes-no questions in GuanM, specifically emphasizing the examination of the boundary tone of the question tune. The research also aims to explore the interaction between lexical tone and tune, with the T1T1 sandhi involved.

2. Methods

2.1. Subjects

Eight native speakers of GuanM (three females; age range: 20 to 45 years old; mean age: 31.88 years old) participated in this study. They were from Weinan, China, and spoke GuanM on a daily basis. None of them had any hearing or speaking disorders.

2.2. Materials and procedure

Fifteen disyllabic words were selected from each of the T1T1, T2T2, and T4T4 sequences, as detailed in Table 1. All selected words were sonorants to avoid the undesirable F0 frequencies typically caused by obstruents, ensuring more accurate alignment and annotation.

During the experiment, speakers were instructed to read only the final disyllabic word from the carrier sentence "This is xx ?" for questions, and "This is xx." for statements, where "xx" represents the stimulus word. The question tune was designed to create a scenario in which speakers sought confirmation.

 10.21437/SpeechProsody.2024-256

Each stimulus was repeated three times. Recordings were made at a sampling rate of 44.1 kHz using Audacity software.

Table 1: *Stimuli used for recordings.*

T1T1	T2T2	T4T4
u31 ia31 "crow"	ma24 iou24 "sesame oil"	laŋ55 man55 "romance"
yan31 iaŋ31 "mandarin duck"	mian24 iaŋ24 "sheep"	mian55 liau55 "fabric"
iau31 ɲiɛ31 "evildoer"	mian24 ma24 "linen fabric"	mian55 mau55 "appearance"
luo31 iɛ31 "fallen leaves"	lan24 mei24 "blueberry"	iaŋ55 mau55 "appearance"
ien31 io31 "music"	iaŋ24 mau24 "wool"	ŋau55 yen55 "Olympic games"
lu31 ien31 "recording"	nuŋ24 mien24 "farmer"	uæ55 mæ55 "takeout"
uan31 yɛ31 "crescent moon"	nan24 men24 "south gate"	miŋ55 yen55 "destiny"
yɛ31 li31 "experience"	iaŋ24 mei24 "bayberry"	uæ55 mau55 "foreign trade"
yɛ31 mo31 "end of month"	liaŋ24 mien24 "good citizen"	miŋ55 liŋ55 "order"
mu31 lu31 "catalogue"	iau24 ɲian24 "rumor"	li55 yŋ55 "use"
iou31 yɛ31 "superior"	mau24 ly24 "donkey"	liŋ55 luei55 "alternative"
mu31 y31 "bathing"	mei24 iou24 "kerosene"	ŋæ55 mei55 "ambiguous"
ian31 iɛ31 "tobacco"	iou24 lyen24 "cruise"	miŋ55 ŋan55 "homicide"
iŋ31 liɛ31 "heroes"	mau24 ɲiou24 "yak"	i55 lyen55 "discussion"
iau31 ɲio31 "invite"	iou24 y24 "squid"	i55 mæ55 "charity sale"

2.3. Measurement and statistical analyses

We took acoustic measurements including duration, mean F0, maximum and minimum F0 and F0 range (the difference between maximum and minimum F0) at both the syllable level and the entire utterance level. For the data analysis, we fitted linear mixed-effects models in *R* [15]. *Tone* (T1T1, T2T2, and T4T4) and *Tune* (Question and Statement), along with their interaction, were treated as fixed effects. At the syllable level, *Syllable* was also included as a fixed effect. The random effects included *Subject*, *Item*, and *Repetition*, while *Duration* and pitch parameters were the dependent variables. Furthermore, we conducted post-hoc pairwise comparisons of significant effects.

3. Results

3.1. Annotation and alignment

We considered only two of the three repetitions of each recording from each speaker, yielding a total of 180 recordings per speaker (3 tones * 15 disyllables * 2 tunes * 2 repetitions). Each individual spectrogram was visually inspected, followed by annotation in *Praat* [16]. To ensure consistency in annotation, the onset of the utterances was marked based on the start of F0 and the presence of the F1, F2, and F3 formants. The offset was determined by the disappearance of F0, F1, F2, and F3 formants. In Figure 2, waveforms of one stimulus from T1T1, from a speaker, are presented, featuring both a statement (on the left) and a question (on the right), with their spectrograms illustrating the F0 contours.

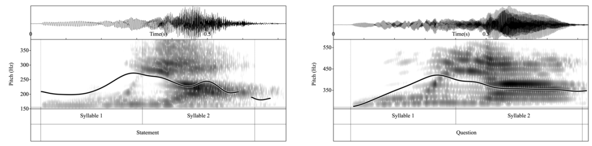

Figure 2: *Statement tune (left) and question tune (right) from the same subject.*

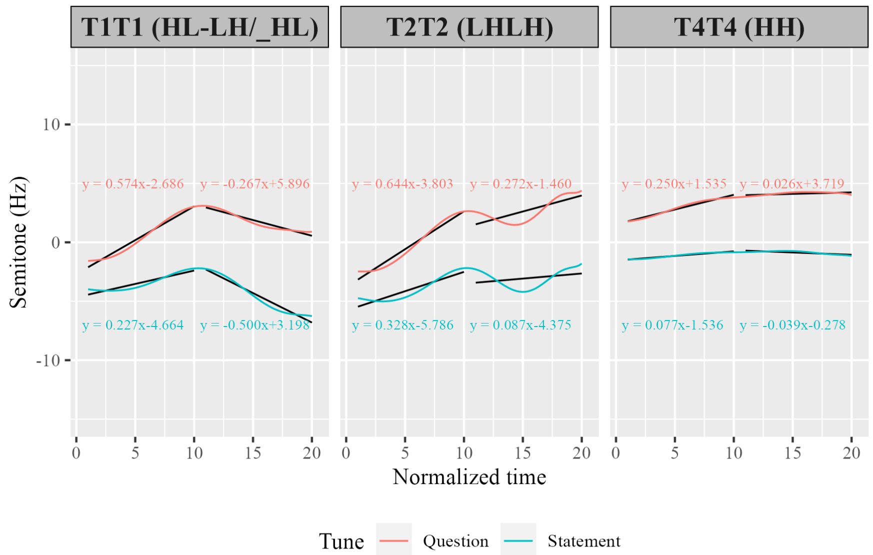

Figure 1: *Average normalized representation of F0 curves of disyllabic words with their linear regressions.*

3.2. Normalized representation of F0 curves

First, we extracted 10 points of F0 of each syllable and subsequently normalized F0 to a semitone scale based on the semitone transformation using each speaker's average F0 value in their entire utterances, plotted in Figure 1. Upon the direct observation of the figure (smoothed lines), overall, a consistent pattern emerged across all three tones: the F0 curves of the question tune consistently were above those of the statement tune, indicating a higher F0 of the question tune.

Moreover, for the question tune of T1T1, the first syllable maintained its sandhi contour, changing from a falling tone to a rising tone and the second syllable did not exhibit as significant a drop as observed in its statement counterpart. Moving to T2T2, the F0 curve of the second syllable of the question tune had an obvious rise compared to its statement, suggesting a potential function of a high boundary tone. For T4T4, the F0 curve in the question tune ascended even further, emphasizing a higher pitch associated with the interrogative tune.

3.3. F0 change rate

The linear regression analysis was conducted using the normalized time point as the independent variable and the semitone value as the dependent variable for each syllable of each tone with both tunes:

$$y = kx + b \quad (1)$$

where x is the time point; y is the semitone value; k is the F0 slope, and b is the intercept. This approach allowed us to assess how the pitch changed over time within each syllable.

As shown in Figure 1 and Table 2, for the first syllable across all three tones, the slopes were consistently positive in both the question and statement tunes, indicating an overall rising trend in pitch. Particularly, the slopes of the question tune were larger than those for the statement tune, which indicates a faster rate of pitch change in questions compared to statements. This suggests that questions might use a more dynamic intonation pattern. In contrast, for the second syllable in T1T1, the slope was negative for both tunes, highlighting a falling pitch trend. In this case, the statement tune exhibited a strong downward pitch movement (slope of -0.500), compared to the question tune (slope of -0.267). However, for the tones T2T2 and T4T4, the slopes for questions in the second syllable remained larger than those for statements, reinforcing the observation that pitch changes more rapidly in questions.

Table 2: *Intercepts and slopes.*

Tone	Tune	Syllable	Slope	Intercept
T1T1	Q	Sy.1	0.574	-2.686
		Sy.2	-0.267	5.896
	S	Sy.1	0.227	-4.664
		Sy.2	-0.500	3.198
T2T2	Q	Sy.1	0.644	-3.803
		Sy.2	0.272	-1.460
	S	Sy.1	0.328	-5.786
		Sy.2	0.087	-4.375
T4T4	Q	Sy.1	0.250	1.535
		Sy.2	0.026	3.719
	S	Sy.1	0.077	-1.536
		Sy.2	-0.039	-0.278

3.4. Duration

At the whole utterance level, the duration of the question tune was generally longer than that of the statement tune. Specifically, the duration was significantly longer for T1T1 ($X^2(1) = 299.924$, $p < 0.001$) and T4T4 ($X^2(1) = 22.308$, $p < 0.001$). However, this difference was not statistically significant for T2T2 ($X^2(1) = 1.358$, $p = 0.244$).

However, at the syllable level, a consistent pattern was observed that the duration of the first syllable in the question tune was generally shorter than the second syllable. This difference was statistically significant for T2T2 and T4T4 ($X^2(1) = 6.902$, $p = 0.017$; $X^2(1) = 81.001$, $p = 0.004$ respectively; but not for T1T1: $X^2(1) = 1.236$, $p = 0.266$). Conversely, the second syllable of the question tune was significantly longer than its statement counterpart across all three tones (T1T1: $X^2(1) = 499.057$, $p < 0.001$; T2T2: $X^2(1) = 19.423$, $p < 0.001$; T4T4: $X^2(1) = 81.001$, $p < 0.001$). This suggests a potential lengthening effect induced by the boundary tone of the question tune.

3.5. Register

Mean F0, maximum F0, and minimum F0 had significantly higher values in the question tune compared to its statement counterpart, observed at both the syllable and whole utterance levels, shown in Table 3, indicating that the register of the question tune was higher and lifted.

Table 3: *Mean, maximum and minimum F0 of three tones (Hz).*

Tone	Tune	Sy.	Mean F0	Max F0	Min F0
T1T1	Q	Sy.1	213.207	246.89	173.273
		Sy.2	225.562	249.204	205.058
	S	Sy.1	165.992	188.264	144.623
		Sy.2	160.347	185.806	131.226
T2T2	Q	Sy.1	203.607	235.884	164.677
		Sy.2	236.279	266.259	207.895
	S	Sy.1	160.254	179.865	141.561
		Sy.2	167.859	186.262	151.921
T4T4	Q	Sy.1	237.601	253.668	211.63
		Sy.2	253.414	262.006	239.053
	S	Sy.1	186.374	197.544	173.409
		Sy.2	188.158	193.444	180.565

3.6. F0 range

At the whole utterance level, the F0 range of the question tune demonstrated a significant increase across three tones (T1T1: $X^2(1) = 13.165$, $p < 0.001$; T2T2: $X^2(1) = 250.710$, $p < 0.001$; T4T4: $X^2(1) = 55.085$, $p < 0.001$). However, a closer examination at the syllable level revealed that this trend persisted for T2T2 and T4T4 in both syllables. In contrast, for T1T1, the second syllable of the question tune had a significantly smaller F0 range ($X^2(1) = 15.046$, $p < 0.001$) compared to its statement counterpart, while the first syllable had a significantly larger F0 range ($X^2(1) = 124.148$, $p < 0.001$), potentially influenced by the presence of a high boundary tone in the second syllable that mitigated F0 expansion in the question tune.

4. Discussion

4.1. Register

The question tune has a larger intercept than the statement tune (shown in Table 2) across all tones and syllables, indicating a higher baseline F0 for questions than for statements. Additionally, the normalized presentation, which depicts F0 curves for questions consistently above those for statements across all three tones, along with statistical findings of higher mean F0 and maxima for the question tune at both the syllable and whole phrase levels, confirms a higher and a raised register for the question tune in GuanM.

4.2. High boundary tone (H%)

The normalized presentation of F0 curves reveals distinct patterns: In T1T1, the second syllable does not fall as much in questions compared to statements, while in T2T2, the second syllable shows a greater increase in questions than in statements. This observation is supported by F0 change rate analysis, which indicates a smaller absolute value of the F0 slope for the second syllable in T1T1 questions and a higher F0 slope in T2T2 questions. This suggests the presence of a high boundary tone in the question tune that moderates the falling tone, preventing it from decreasing as much as in its statement counterpart, and enhances the rising tone, allowing it to increase more significantly. Furthermore, according to Pierrehumbert [17], an intonational phrase boundary occurs either when there is no hesitation pause or when the final syllable of a phrase is lengthened. The high boundary tone from the question tune causes the lengthening of the second syllable, which is also the last syllable of the intonational phrase, in GuanM.

Lin [18] proposed that the "boundary tone" exclusively distinguishes questions from statements in Chinese. However, the argument here posits that in GuanM, both the raised register and the boundary tone contribute to the question intonation. This interaction, along with the lexical tone, prevents falling tones from dropping excessively and assists rising tones in increasing.

5. Conclusions

This study investigates the syntactically unmarked question tune in GuanM. We propose that its structure comprises maintaining the tone sandhi contour, a raised register, and a high boundary tone (H%).

The current experiment focuses on the production of the question tune by utilizing only sentence-final disyllabic words carrying the intonational phrase boundary tone information. To gain a more comprehensive understanding, it would be necessary to extend the investigation to longer utterances with more boundaries within the prosodic hierarchy.

6. References

[1] Q. Gong, "Guanzhong fangyan de biandiao he bianyin [The tone and sound change of Guanzhong dialect]," *Yuwen yanjiu,* vol. 4, pp. 55-58, 1995.

[2] E. Dornisch, Multiple-wh-questions in Polish: The Interactions between Wh-phrases and Clitics, ProQuest Dissertations and Theses, 1998.

[3] N. Rojina, "The acquisition of wh-questions in Russian," *Nordlyd,* vol. 32, no. 1, p. 68–87, 2004.

[4] M. Durrell, Hammer's German grammar and usage, 5 ed., Abingdon: Routledge, 2011.

[5] J. M. Dewaele, "Word order variation in French interrogative structures," *ITL-International Journal of Applied Linguistics,* vol. 125, no. 1, pp. 161-180, 1999.

[6] Y. R. Chao, A Grammar of Spoken Chinese, Berkeley and Los Angeles: University of California Press, 1968.

[7] M. Tsuchihashi, "The speech act continuum: An investigation of Japanese sentence final particles," *Journal of Pragmatics,* vol. 7, no. 4, pp. 361-387, 1983.

[8] J. Vaissière, "Perception of intonation," in *The handbook of speech perception*, Malden ; Oxford, Blackwell, 2005, pp. 236-263.

[9] D. Bolinger, "Intonation across languages," *Universals of human language,* vol. 2, pp. 471-524, 1978.

[10] A. Rialland, "Question prosody : an African perspective," in *Tones and Tunes. Volume 1, Typological Studies in Word and Sentence Prosody*, Berlin ; Boston , De Gruyter Mouton, 2007, pp. 35-62.

[11] A. V. Lyovin, "Reviewed Work: Ton i intonacija v sovremennom kitajskom jazyke (Tone and intonation in Modern Chinese) by Mixail Kuz'mič Rumjancev," *Journal of Chinese Linguistics,* vol. 6, no. 1, pp. 120-168, 1978.

[12] X.-n. S. Shen, The prosody of Mandarin Chinese, vol. 118, Berkeley: University of California Press, 1989.

[13] A. T. Ho, "Intonation variation in a Mandarin sentence for three expressions: interrogative, exclamatory and declarative," *Phonetica,* vol. 34, no. 6, pp. 446-457, 1977.

[14] F. Liu and Y. Xu, "Parallel encoding of focus and interrogative meaning in Mandarin intonation. , 62(2-4), 70-87.," *Phonetica,* vol. 62, no. 2-4, pp. 70-87, 2005.

[15] R. C. Team, "R: A Language and Environment for Statistical Computing," R Foundation for Statistical Computing, Vienna, Austria, 2023. [Online]. Available: https://www.R-project.org/.

[16] P. Boersma and D. Weenink, "Praat: doing phonetics by computer. Version 6.3.16," 29 August 2023. [Online]. Available: from http://www.praat.org/.

[17] J. B. Pierrehumbert, The phonology and phonetics of English intonation, (Doctoral dissertation, Massachusetts Institute of Technology), 1980.

[18] M. Lin, "Yiwen he chenshu yuqi yu bianjiediao [Interrogative mood and boundary tone in Chinese]," *Zhongguo Yuwen,* vol. 4, pp. 364-379, 2006.