2024. *Proc Ling Soc Amer* 9(1). 5676. https://doi.org/10.3765/plsa.v9i1.5676.

Perception of the question tune in Guanzhong Mandarin

Jiarui Zhang*

Abstract. This study investigates the perception of the question tune in Guanzhong Mandarin and the interaction between tone and tune using the AX paradigm. Our findings reveal a complex interplay between lexical tones and question tunes, in which when the lexical tone is rising, it is more difficult to perceive question tunes. Furthermore, the research argues that the question tune retrieval requires extra working memory load and cognitive processing, because of the tune information brought by a high register and a high boundary tone.

Keywords. Chinese dialect; tone and intonation interaction; perception of tune; AX task; prosody

1. Introduction. Guanzhong Mandarin (hereinafter referred to as GuanM), a sub-dialect of Mandarin spoken in Xi'an and its neighboring cities in China, shares many similarities with Standard Mandarin in terms of syntactic structures and prosodic features. GuanM has four lexical tones: T1T1 (31, HL_{lr} Tone), T2T2 (24, LH Tone), T3T3 (53, HL_{hr} Tone). T4T4 (55, H Tone). The tone sandhi rules in GuanM are:

(1) a. T1 + T1 → T2 + T1: HL → LH (24) /__ HL_{lr} e.g: [k^hæ31→24 xua31] "blossom"
b. T3 + T3 → T1 + T3: HL_{hr} → HL_{lr} (31) /__ HL_{hr} e.g: [ʂʏu53→31 piaw53] "watch"

In (1)a, the first falling tone, T1, changes to a rising tone, T2, when it precedes another T1. In (1)b, the first falling tone, T3, which has a high tone register, lowers its register to T1 before another T3.

It is argued that the tune for intonational yes-no questions in GuanM features a high boundary tone and a higher register (Zhang, 2024). The intonational tune of unmarked yes-no questions maintains the tone sandhi contour if there are any, and the high boundary tone does not change the contour shape of the last syllable or last prosodic phrase. This high boundary tone acts in a way that, for the last syllable, it prevents the falling tone from decreasing further and slows down any pitch changes and it facilitates an increase in the rising tone and accelerates its upward movement.

While previous research has showed challenges in perceiving question tunes associated with rising tones across Chinese dialects, those studies primarily emphasized tone perception and utilized different methodologies. In response, our research used the AX discrimination task (Dupoux et al. 1997).

2. Research questions. Does the perception of a question tune become more challenging with a rising tone, due to the interaction between the lexical tone and the intonation, particularly when considering how the shapes of the lexical tone and question tune contours correspond? If interactions exist, does the order of tune retrieval during the AX task influence this interaction? What can this tell us?

* We thank the audience at 2024 Annual Meeting of the Linguistic Society of America for comments and suggestions. Author: Jiarui Zhang (jiarui.zhang@ling-phil.ox.ac.uk), University of Oxford.

3. Methodology. Forty subjects from Weinan, China (20 females, mean age: 36.95 years) participated in the experiment. They had no reported hearing or speaking disorders and gave their consent before participating in the experiment.

3.1. STIMULI AND PROCEDURE. For the experiment, 36 disyllabic words were selected, with 12 each from T1T1, T2T2, and T4T4 (as shown in Table 1). These words were recorded by a 24-year-old female native speaker of GuanM, using both question and statement intonations. The words were placed at the end of the sentence "This is xx?" to elicit the question tune and "This is xx." for the statement tune, where "xx" represents the disyllabic audio stimulus.

T1T1	**T2T2**	**T4T4**
u31 ia31 "crow"	ma24 iou24 "sesame oil"	laŋ55 man55 "romance"
yan31 iaŋ31 "mandarin duck"	mian24 iaŋ24 "sheep"	mian55 liau55 "fabric"
iau31 ŋiɛ31 "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 ŋian24 "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 ŋiou24 "yak"	i55 lyen55 "discussion"
iau31 ŋio31 "invite"	iou24 y24 "squid"	i55 mæ55 "charity sale"

Table 1. Stimuli list used in the study

The study used an AX paradigm with four combinations. In each combination, the first sound (A) and the second sound (X) consisted of the same word, produced with either a state-ment or a question tune. The combinations were as follows: A= Statement, X= Statement (SS); A= Statement, X= Question (SQ); A= Question, X= Question (QQ); A= Question, X= Statement (QS). These sequences were presented twice to each subject, with the order of presentation ran-domized to control for potential order effects and ensure reliability in the responses. In total there were 288 trials per subject (3 Tones x 12 disyllabic words x 4 combinations x 2 repetitions). The experiment was conducted in PsychoPy (Peirce et al. 2019). Subjects were seated in a quiet room of a local community and wore headphones. They were instructed to listen carefully to each pair of sounds and to determine whether the tune of the second sound (X) matched that of the first one (A) by pressing the key "1" for "same" and "0" for "different" as quickly and accu-rately as possible. Subjects were given 2000ms to respond. Between the two sounds was a 200ms interval.

4. Measurement and statistical analyses. To assess the subjects' ability to discriminate be-tween different pairs, we calculated *d*-prime (*d*') using the framework of signal detection theory (SDT), as described by Macmillan and Creelman (2004). The calculation of *d*-prime was carried out for each subject across each tone, divided into two pairs based on the tune of the first sound (A): SS_SQ (when A = Statement) and QQ_QS (when A = Question), using the formula: $d' = Z(\text{hit rate}) - Z(\text{false alarm rate})$. Here, the hit rate is the proportion of times subjects correctly identified the second sound (X) as different when it was, in fact, different (for SQ or QS). The false alarm rate is the proportion of times subjects incorrectly identified the second sound as dif-ferent when it was actually the same (for SS or QQ). *d*-prime score is the difference between these *Z*-score transformations. To analyze the *d*-prime scores, we fitted a linear mixed-effects model using the packages *LME4* (Bates et al. 2015) in *R* (R core team 2023), where *d*-prime score was the dependent variable, with *Pair* (SS_SQ and QQ_QS) and *Tone* (T1T1, T2T2 and T4T4) and their interactions as fixed effects, and *Subject* as a random effect.

We also measured reaction times (RTs) and accuracy. For the RT analysis, we fitted a linear mixed-effects model. In this model, RT was the dependent variable, with *Tone* (T1T1, T2T2 and T4T4), *AX Sequence* (SS, SQ, QQ and QS) and their interactions as fixed effects and *Subject*, *Stimulus Item* and *Repetition* as random effects. The accuracy data was analyzed with a general-ized linear mixed-effects model designed for a binomial distribution (correct vs. incorrect responses). Both fixed and random effects in this model were the same as those in the RT model. We further conducted post-hoc pairwise comparisons for significant effects using Tukey's HSD tests in the *emmeans* package in R (Lenth et al. 2018).

5. Results. Before the statistical analyses, we conducted an outlier removal of RTs, where RTs that were more than 3 standard deviations above or below each subject's mean RT were identified as outliers and excluded, resulting in a data loss of 1.48% (170 trials).

5.1. *D*-PRIME SCORES. In this study, the *d*-prime scores with respect to the *Pair* and *Tone* were plotted in Figure 1.[1] The statistical results showed a significant difference in *d*-prime scores across tones. Specifically, both T2T2 and T4T4 had significantly lower *d*-prime scores compared to T1T1, with differences of 0.74 and 0.31, respectively, both of which were statistically signifi-cant ($p < 0.001$ for both). However, for the factor *Pair*, there was no significant effect found ($p = 0.474$; see Table 2 for more details).

[1] The error bars were included to show the standard error of the mean by Pairs and Tones.

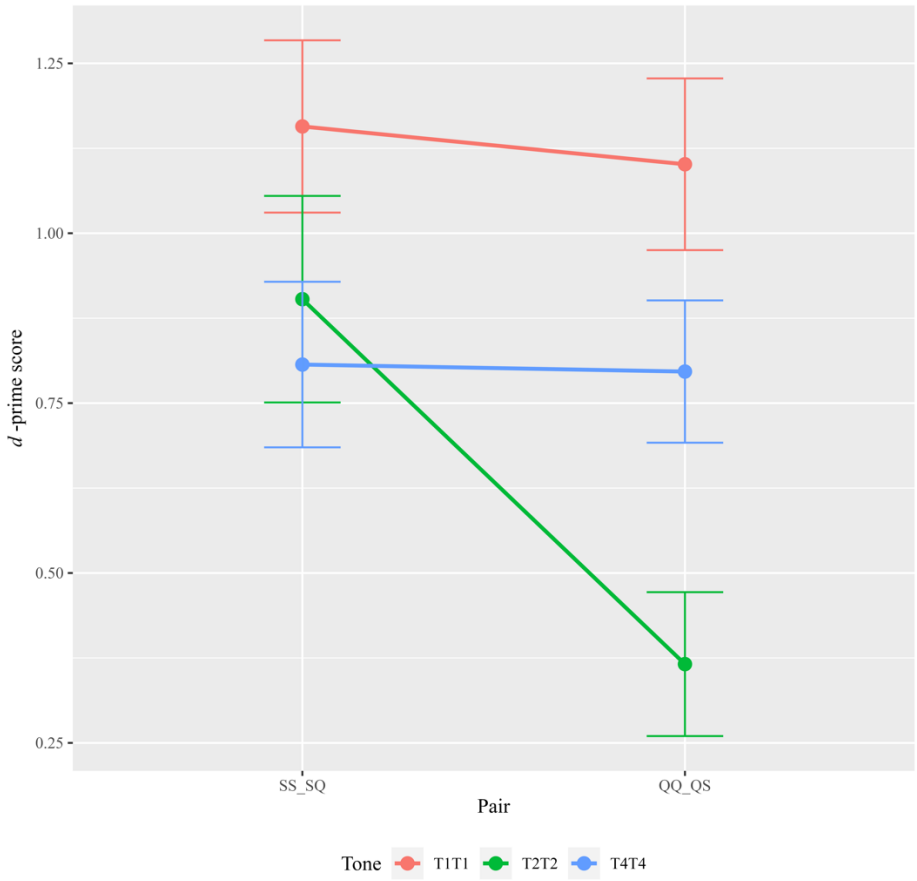

Figure 1. D-prime scores for T1T1, T2T2 and T4T4 by Pair SS_SQ and QQ_QS

Predictors	*Estimates*	*CI*	*p*
(Intercept)	1.10	0.86 – 1.35	**<0.001**
Tone [T2T2]	-0.74	-0.89 – -0.58	**<0.001**
Tone [T4T4]	-0.31	-0.46 – -0.15	**<0.001**
Pair [SS_SQ]	0.06	-0.10 – 0.21	0.474
Tone [T2T2] × Pair [SS_SQ]	0.48	0.27 – 0.70	**<0.001**
Tone [T4T4] × Pair [SS_SQ]	-0.05	-0.26 – 0.17	0.680

Table 2. Model output for d-prime scores. Square brackets indicate the factor level comparisons.

Post-hoc analysis further revealed that *d*-prime score for T2T2 was the lowest when the first sound (A) had a question tune. This score was significantly lower than that of other pairs within the interaction. Details of these statistical results can be found in Table 3.

Additionally, the analysis showed no significant difference in *d*-prime scores for T1T1 and T4T4 when comparing SS_SQ and QQ_QS pairs ($p = 0.9796$ and $p = 1.0000$, respectively), suggesting that the ability to discriminate between tunes did not vary whether the first sound had a question or a statement tune. However, within the QQ_QS pair (where A was a question tune), the *d*-prime score for T1T1 was significantly higher than T2T2 by 0.7356 ($p < 0.001$) and sig-

nificantly higher than T4T4 by 0.3050 ($p = 0.0016$), indicating a higher discrimination rate for T1T1 compared to T2T2 and T4T4. Furthermore, the *d*-prime score for T4T4 was significantly larger than that for T2T2 by 0.4305 ($p < 0.001$), suggesting a gradation in discrimination ability with the strongest being for T1T1, followed by T4T4, and T2T2 being the least.

Within the SS_SQ pair (when A was a statement tune), the *d*-prime score for T1T1 was significantly higher than T2T2 by 0.2542 ($p = 0.0155$), and higher than T4T4 by 0.3504 ($p = 0.0002$). No significant difference was observed between T2T2 and T4T4 ($p = 0.8165$), which suggested when the first sound was a statement tune, subjects had the highest dsicrimintation rate for the tune difference in T1T1 with no difference in T2T2 and T4T4.

contrast	estimate	SE	df	t.ratio	p.value
T2T2 QQ_QS - T1T1 SS_SQ	-0.7912	0.0776	195	-10.200	<.0001
T2T2 QQ_QS - T1T1 QQ_QS	-0.7356	0.0776	195	-9.483	<.0001
T2T2 QQ_QS - T2T2 SS_SQ	-0.5370	0.0776	195	-6.923	<.0001
T2T2 QQ_QS - T4T4 SS_SQ	-0.4408	0.0776	195	-5.683	<.0001
T2T2 QQ_QS - T4T4 QQ_QS	-0.4305	0.0776	195	-5.550	<.0001

Table 3. Pairwise post-hoc comparisons for T2T2 within the QQ_QS pair

5.2. REACTION TIMES. There were no significant differences for *Tone* or *AX sequence*, with the exception of the QS sequence. In this sequence, reaction times were significantly longer ($p = 0.038$), indicating that subjects required more time to react when recalling a question intonation to match it with a statement intonation. See Table 4 and Table 5 below for the detailed statistical outcomes.

AX Sequence	Tone	N	Mean	SD
A = Statement, X = Statement	T1T1	759	369.547	293.189
	T2T2	737	364.314	307.87
	T4T4	705	375.063	316.729
A = Statement, X = Question	T1T1	701	346.931	314.606
	T2T2	616	368.115	314.289
	T4T4	615	376.104	301.112
A = Question, X = Question	T1T1	758	364.201	288.411
	T2T2	663	395.002	329.39
	T4T4	747	375.593	302.83
A = Question, X = Statement	T1T1	701	389.085	324.121
	T2T2	478	375.90	299.69
	T4T4	543	389.00	311.12

Table 4. Reaction times of correct responses across different AX sequences and tones

Predictors	*Estimates*	*CI*	*p*
(Intercept)	360.38	310.30 – 410.45	**<0.001**
AX Sequence [A=S, X=Q]	-14.58	-43.41 – 14.26	0.322
AX Sequence [A=Q, X=Q]	-5.60	-33.80 – 22.59	0.697
AX Sequence [A=Q, X=S]	30.54	1.69 – 59.39	**0.038**
Tone [T2T2]	-11.36	-41.08 – 18.36	0.454
Tone [T4T4]	3.51	-26.52 – 33.53	0.819
AX Sequence [A=S, X=Q] × Tone [T2T2]	35.46	-6.11 – 77.02	0.095
AX Sequence [A=Q, X=Q] × Tone [T2T2]	34.65	-6.10 – 75.41	0.096
AX Sequence [A=Q, X=S] × Tone [T2T2]	9.70	-33.57 – 52.97	0.660
AX Sequence [A=S, X=Q] × Tone [T4T4]	32.41	-9.38 – 74.20	0.129
AX Sequence [A=Q, X=Q] × Tone [T4T4]	6.06	-34.26 – 46.39	0.768
AX Sequence [A=Q, X=S] × Tone [T4T4]	3.85	-38.73 – 46.42	0.859

Table 5. Model output for RTs. Square brackets indicate the factor level comparisons

5.3. ACCURACY. It is directly observed from Figure 2 that accuracy for T2T2 in the sequence of QS was particularly low, at only 50.8%. This implies that the subjects' responses in this sequence were more like a chance-level decision-making. In contrast, for T1T1, accuracy rates across all four sequences were relatively high. Sequences with matching A and X tunes (SS and QQ), demonstrated overall higher accuracy than those with differing tunes (SQ and QS). Moreover, the sequence of QS exhibited the lowest accuracy rate across three tones.

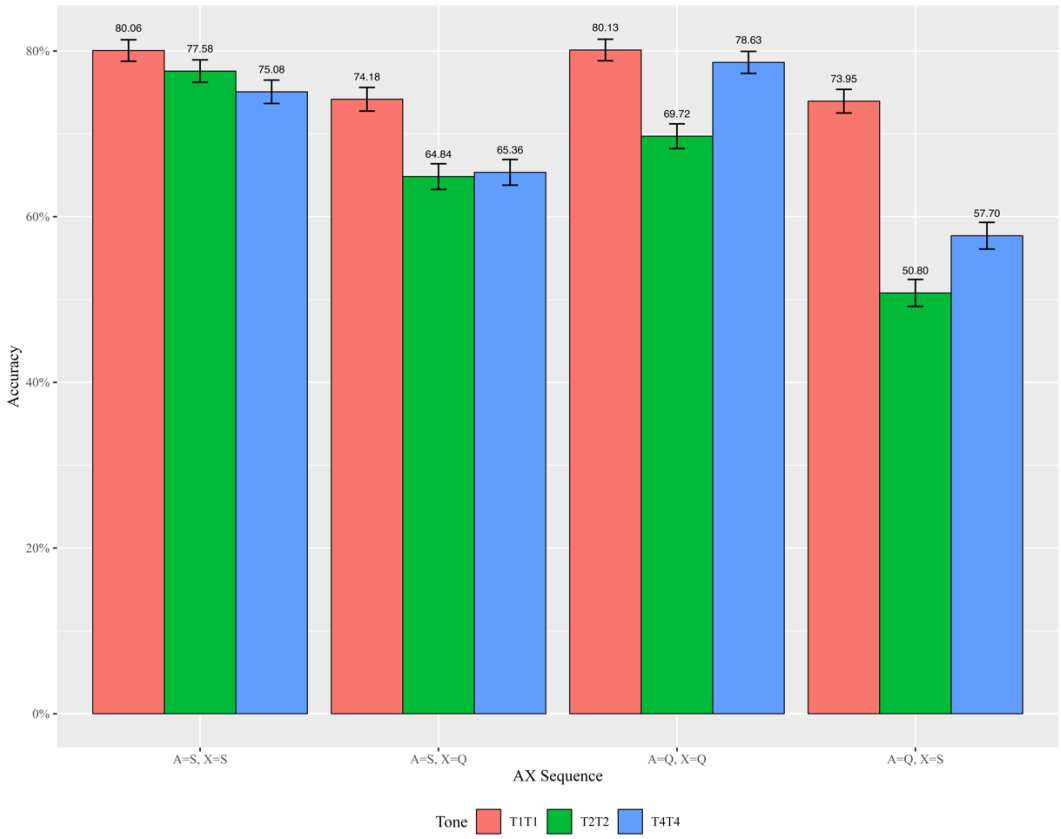

Figure 2. Accuracy across different AX sequences and tones

The statistical analysis showed that for AX sequences where the tunes of A and X did not match (SQ and QS), the odds ratios (OR) were both 0.67, indicating a significant 33% reduction in the odds of responding accurately to these sequences, with p-values of 0.001 for both. Further analysis of the interaction between *AX Sequence* and *Tone*, especially the interaction between the QS sequence and T2T2, showed the OR decreased to 0.34, which represents a substantial 66% decrease in the odds of accuracy with a strong statistical significance ($p < 0.001$). While other interactions, like the SQ sequence paired with T2T2, also showed lower odds of accuracy, the drop was less dramatically. See Table 6 for more interaction details.

Moving to the post-hoc analysis, which focused on the AX sequence comparisons within each tone- where the tunes of A and X either matched (SS and QQ) or did not match (SQ and QS), several key points were found. For sequences where A and X tunes were the same (SS and QQ), no significant difference in accuracy rate was found for T1T1 and for T4T4 ($p = 1$ and $p = 0.7471$, respectively). This suggested for T1T1 and T4T4, the tune did not influence accuracy rates; same tunes in a sequence were generally easier to identify. However, for T2T2 in the matched sequence, the SS sequence showed a significantly higher log-odds of accuracy compared to QQ, with an estimated difference of 0.484 ($p = 0.0014$), indicating identifying QQ in T2T2 was more challenging than SS. Regarding sequences where A and X tunes were different (SQ and QS), the accuracy rate for T1T1 showed no significant difference (p = 1), suggesting that the order between the statement and the question tune or vice versa did not affect accuracy for this tone. However, for T2T2 and T4T4, the SQ sequence was significantly more accurate than QS, with log-odds of difference of 0.706 ($p < 0.0001$) and 0.393 ($p = 0.0105$), respectively.

This implied that for these tones, subjects were more accurate in discriminating from statement to question but not from question to statement.

Predictors	*Odds Ratios*	*CI*	*p*
(Intercept)	5.36	3.65 – 7.87	**<0.001**
AX Sequence [A=S, X=Q]	0.67	0.53 – 0.85	**0.001**
AX Sequence [A=Q, X=Q]	1.01	0.79 – 1.29	0.938
AX Sequence [A=Q, X=S]	0.67	0.53 – 0.84	**0.001**
Tone [T2T2]	0.84	0.62 – 1.15	0.282
Tone [T4T4]	0.72	0.53 – 0.98	**0.038**
AX Sequence [A=S, X=Q] × Tone [T2T2]	0.69	0.50 – 0.96	**0.025**
AX Sequence [A=Q, X=Q] × Tone [T2T2]	0.61	0.44 – 0.85	**0.003**
AX Sequence [A=Q, X=S] × Tone [T2T2]	0.34	0.25 – 0.47	**<0.001**
AX Sequence [A=S, X=Q] × Tone [T4T4]	0.84	0.61 – 1.16	0.284
AX Sequence [A=Q, X=Q] × Tone [T4T4]	1.24	0.89 – 1.74	0.203
AX Sequence [A=Q, X=S] × Tone [T4T4]	0.57	0.42 – 0.79	**0.001**

Table 6. Model output for Accuracy. Square brackets indicate the factor level comparisons.

6. Discussion and conclusions. The finding that reaction times were longest for the QS sequence indicates that participants required more time to process, suggesting that the task retrieving the question tune and matching the subsequent information of the statement tune is cognitively demanding.

Concerning the tone interaction during the tune perception, for T1T1, a falling tone, there is a conflict with the question tune in terms of lexical contour shapes. Despite the imposition of the question tune, T1T1still maintains its lexical contour shape, which aligns with the definition of a high boundary tone in GuanM that does not change the lexical contour. As a result, the accuracy rate remains consistently high across all four sequences (SS, SQ, QQ and QS) for T1T1, and the *d*-prime scores are the highest compared to other tones in both SS_SQ and QS_SQ pairs. With T4T4, a high-level tone, the *d*-prime scores did not show any statistical differences between SS_SQ and QS_SQ pairs, indicating equivalent discrimination abilities regardless of the preceding tune. Nonetheless, the accuracy rate was low for QS and SQ sequences, suggesting that pairing a high-level tone with a statement tune with a question tune marked by a high boundary tone can introduce perceptual difficulties. For T2T2, the *d*-prime scores were significantly lower for the QQ_QS pair than for the SS_SQ pair, emphasizing a reduction in the ability to discriminate the tune differences in the QQ_QS where the first sound had a question tune. This suggests that combining a question tune with a rising tone showed significant challenges in perception, as reflected in the lowest accuracy scores for the QS sequence, indicating substantial difficulty in the perception of T2T2 with a question tune.

Moreover, we argue that when a question tune is initially activated and stored in short-term memory, its retrieval is challenging. This complexity arises from the necessity of accessing additional information about the tune, specifically its register and contour shape. During this retrieval process, the lexical tone interacts with the tune. Specifically, when the contour shape of the lexical tone does not match the tune contour, retrieval is simple; however, when the contour shapes align, as observed in T2T2, retrieval and perception become significantly more challenging. Furthermore, the necessity of a high boundary tone in the tune of questions is highlighted by our findings. The sensitivity measure (*d*-prime) does not show a significant difference in T4T4, a high tone, regardless of whether the first sound is a question tune, or a statement tune but did differ in T2T2, a rising tone. Additionally, the accuracy in T4T4 is higher than in T2T2. This suggests that recognizing question tunes is influenced not only by the high register but also crucially by the high boundary tone that enhances the rising tone and accelerates the change, necessitating increased processing effort in such discrimination tasks.

This pilot study aims to explore the perception of yes-no question tunes in GuanM and the interaction between tune and tone using the AX paradigm. The findings indicate that in tune perception, tones interact, and the retrieval of the question tune necessitates a short-term memory load concerning both the high register and the contour shape, facilitated by the high boundary tone.

References

Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. 2015. Fitting linear mixed-effects models using Lme4. *Journal of Statistical Software* 67 (1). 1–48. https://doi.org/10.18637/jss.v067.i01.

Dupoux, Emmanuel, Christophe Pallier, Nuria Sebastian & Jacques Mehler. 1997. A destressing "deafness" in French?. *Journal of Memory and Language* 36. 406–421. https://doi.org/10.1006/jmla.1996.2500.

Macmillan, Neil A. & C. Douglas Creelman. 2004. *Detection theory: A user's guide*. New York: Psychology Press.

Peirce, Jonathan, Jeremy R. Gray, Sol Simpson, Michael MacAskill, Richard Höchenberger, Hiroyuki Sogo, Erik Kastman & Jonas Kristoffer Lindeløv. 2019. PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods* 51. 195–203. 10.3758/s13428-018-01193-y

R Core Team. 2023. *R: A language and environment for statistical computing*. https://www.r-project.org.

Zhang, Jiarui. 2024. Intonation of yes-no questions in Guanzhong Mandarin. Manuscript.