

# The Role of the Sentence-Final Neutral Tone in Mandarin Intonation Perception

Xinyuan Shi

School of Chinese Language and Literature, Nanjing Normal University

Nanjing, Jiangsu, China

E-mail: xinyuan.shi@njnu.edu.cn

Aishu Chen (Corresponding author)

School of English Education, Guangdong University of Foreign Studies

Guangzhou, Guangdong, China

E-mail: aishu\_chen@gdufs.edu.cn

Received: July 21, 2026

Accepted: October 8, 2026

Published: October 24, 2026

doi:10.5296/ijl.v18i5.23939

URL: <https://doi.org/10.5296/ijl.v18i5.23939>

## Abstract

In tone languages like Mandarin Chinese, fundamental frequency (F0) simultaneously encodes lexical tone and intonation. While previous research has established that intonation is primarily carried by sentence-final full-tone syllables, it remains unclear how intonation is perceived when the final position is occupied by a phonologically weak neutral-tone syllable. The present study addresses this question by using SVO sentences with the sentence-final particle *le*, whose surface tonal realization depends on its preceding tone. Thirty native Mandarin listeners completed a two-alternative forced-choice intonation identification task (statement versus question) with three stimulus types: original full sentences, truncated sentences (in which *le* was removed), and switched sentences (in which statement stems were combined with question-final *le* and vice versa). Results of binomial generalized linear mixed-effects models revealed that in the switched condition, listeners categorized utterances based on the intonation carried by the final *le*, regardless of the stem-final lexical tone, demonstrating the dominant role of the sentence-final syllable. In the truncated condition, a mirror-image T2-T4 asymmetry was observed: statement stems showed higher consistency when ending with T2, while question stems showed higher consistency when ending with T4. These findings demonstrate that the neutral-tone final syllable serves as the primary

perceptual carrier of intonation, while the preceding stem provides compensatory cues that are modulated by lexical tonal trajectories. The T2-T4 asymmetry further supports the functional trade-off hypothesis that when lexical tone and intonation trajectories are congruent, their overlap creates perceptual ambiguity in the absence of the sentence-final particle.

**Keywords:** Intonation, Neutral tone, Perception, Mandarin Chinese

## 1. Introduction

Almost all human languages use intonation to convey sentence-level meanings such as sentence type (Gussenhoven, 2004; Ladd, 2008). It is well established that cross-linguistically, higher or rising fundamental frequency (F0) conveys uncertainty and questioning, while lower or falling F0 signals assertiveness and declaration (Bolinger, 1982; Cruttenden, 1981; Gussenhoven, 2002; Ohala, 1983). In tone languages like Mandarin Chinese, F0 information simultaneously encodes word-level lexical meanings and sentence-level pragmatic meanings, giving rise to the interaction between tone and intonation (Chao, 1968). Because of this dual function, intonation identification is particularly challenging and is significantly influenced by the lexical tone of the sentence-final full-tone syllable (Liang & Van Heuven, 2007; B. Xu, 2013; Yuan, 2011). However, the vast majority of previous findings on intonation perception are based on sentence-final syllables bearing full lexical tones. The current study therefore investigates intonation perception when the sentence-final position is occupied by a phonologically weak neutral-tone syllable.

Mandarin Chinese distinguishes lexical meaning through four canonical tones, i.e., T1 (high level), T2 (rising), T3 (dipping), and T4 (falling). In addition to these four tones, the system includes a neutral tone (T0). Whether words containing neutral tones are lexicalized (e.g., *gelge0* “older brother”) or nonlexicalized (e.g., *zhulde0* “pig’s”), the pitch realization of a neutral tone depends on the preceding tonal context, and its duration is shorter than that of the four lexical tones (Tang et al., 2018). Early studies described the surface realization of neutral tone as high-falling after T1 and T2, mid-level after T3, and low-falling after T4 (Cao, 1992), whereas later acoustic research has suggested that neutral tone may have a static mid target (Chen & Xu, 2006). When disyllabic words ending in intrinsic neutral tone (e.g., T0 in *gelzi0* “pigeon”), derived neutral tone (e.g., T0 in *dunlshi0(2)* “stocky”), or full citation tones (e.g., T2 in *qinglchu2* “delete” and T4 in *zhuolzhu4* “get hold of”) occupy the sentence-final position, both neutral tone and T4 were realized with a reduced pitch range at a higher pitch register in questions, whereas T2 was characterized by a slightly expanded pitch range and a higher pitch register (Zhang et al., 2022).

The perception of intonation in Mandarin has been shown to be systematically influenced by the lexical tone of the sentence-final full-tone syllable. Behavioral studies have demonstrated that yes-no questions are easier to identify when sentences end with the falling lexical tone (T4), and more difficult when they end with the rising lexical tone (T2) (Yuan, 2011). This finding indicates that the phonological identity of the sentence-final tone intervenes in the mapping of F0 contours to intonational categories, suggesting that tone and intonation interact at a phonological level. The asymmetry between T2 and T4 in intonation perception

has been further supported by neural evidence. Specifically, the question versus statement contrast elicits a clear mismatch negativity (MMN) for T4-bearing syllables but not for T2-bearing ones, indicating that the question versus statement contrast is more salient on T4 (Ren et al., 2013). In addition, P300 data have shown that statements were easier to identify when the final syllable carried T4 than when it carried T2, while there was no tone-specific asymmetry in question intonation perception (Liu et al., 2016). When the utterance-final syllable carries a neutral tone rather than a full lexical tone, intonation identification is more accurate for neutral tone than for T2, and similarly accurate for neutral tone and T4, regardless of whether the neutral tone was intrinsic or derived (Zhang et al., 2022).

The claim that the sentence-final position of Mandarin sentences carries the majority of intonational cues has been supported by perceptual evidence. For echo questions without interrogative particles ending with a full-tone syllable, removing the final full-tone syllable substantially affected question identification (Lin, 2006). Furthermore, cross-splicing the final syllables caused listeners to recognize the intonation by relying on the final syllable rather than on the pre-final context (Jiang, 2010; Lin, 2012). Liang and Van Heuven (2007) manipulated the overall pitch level of sentences and the pitch level of the final tone independently, finding that the final rise had a much stronger effect on intonation perception than the global pitch level of the entire sentence. This result demonstrates that listeners assign greater perceptual weight to the F0 characteristics of the final syllable than to those of the preceding portion when making intonation judgments.

The current literature leaves two gaps regarding the perceptual mechanisms of Mandarin intonation. First, the perceptual distribution of intonational cues in sentences ending with a neutral-tone syllable remains largely unexplored. While previous research has established that a sentence-final full-tone syllable carries the primary intonational cue, it is unclear what happens when the final syllable is phonologically weak. Specifically, it remains to be determined whether the primary intonational cue is localized on the final neutral-tone syllable, shifted leftward to the pre-final sentence stem, or distributed across both domains. Second, the specific interaction between the stem-final lexical tone and intonation perception in these contexts is not well understood. Because the surface F0 contour of a neutral tone is dependent on the preceding lexical tone, the phonological contrast between the preceding tones may affect how listeners extract sentence-level meaning.

The present study addresses these gaps through a perception experiment using truncated and cross-spliced materials. We propose two main hypotheses. First, because the neutral-tone syllable lacks underlying lexical tone specification, it offers an environment relatively free of tone-intonation conflict at the phonological level. We therefore hypothesize that this final neutral-tone syllable will act as the primary perceptual anchor, carrying the majority of intonational cues, while the preceding sentence stem contributes only secondary cues. Second, we hypothesize that the reliance on the final neutral-tone syllable is systematically modulated by the preceding lexical tone. Specifically, the removal of the final neutral-tone syllable will lead to a perceptual asymmetry between T2 and T4. Drawing on the documented T2-T4 asymmetry in Mandarin intonation perception, we predict that when the final neutral-tone syllable is removed, question intonation will be more likely to be identified if the preceding

stem ends with T4 rather than T2.

## 2. Method

### 2.1 Participants

Thirty native Mandarin-speaking undergraduates and postgraduates aged from 19 to 26 years (16 females,  $M=21.17$  years,  $SD=1.53$  years) participated in the perception experiment. All reported using Mandarin Chinese as their primary language in daily communication. No participant had any neurological or physical impairment, or any speech or hearing impairment. Participants were informed about the perceptual task before the experiment, and they were paid after completing the entire task. The experiment was conducted in accordance with the Declaration of Helsinki and was approved by the Institutional Review Board of the first author's affiliated university.

### 2.2 Material

Two female native speakers of Mandarin Chinese with standard pronunciation were recruited to record the experimental materials. We adopted the test sentences from Shi & Li's (2025) production study of Mandarin intonation. The lexical tone of each syllable at each position within the sentences was strictly controlled. Each sentence consisted of a disyllabic subject, a monosyllabic verb, a trisyllabic object, and the neutral-tone sentence-final particle (SFP) *le* (see Table 1). The particle *le* was selected for the following reasons. First, phonologically, *le* represents one of the most typical neutral-tone syllables in Mandarin Chinese. Second, *le* is syntactically independent and primarily functions as an SFP that does not directly encode declarative or interrogative meanings, allowing intonation to serve as a crucial mechanism for distinguishing sentences that share identical word sequences and lexical meanings. Instead of using isolated sentences, we placed both statement and question versions of each sentence within simple conversations (see Table A1 in the Appendix), thereby maximizing the naturalness of the recordings. The two native speakers were required to read the materials in a natural way, avoiding narrow focus or emotional prosody. All audio recordings were saved as 44100 Hz WAV files, noise-reduced in Audacity, and amplitude-normalized in Praat (Boersma & Weenink, 2023).

Table 1. Neutral tone in the final position of four recording sentences

| No. | Sentence (Pinyin form)                   | Translation (statement version)       |
|-----|------------------------------------------|---------------------------------------|
| 1   | <i>zhang1hao4 zuo4 kao3rou4bing3 le0</i> | ‘Zhanghao made grilled meat pie.’     |
| 2   | <i>wang2bao3 chang2 dun4yu2tou2 le0</i>  | ‘Wangbao tasted fish head stew.’      |
| 3   | <i>li3chao1 dian3 zha2ji1chi4 le0</i>    | ‘Lichao ordered fried chicken wings.’ |
| 4   | <i>zhao4hao2 chi1 shao1sun3gan1 le0</i>  | ‘Zhaohao ate cooked bamboo shoots.’   |

*Note.* The four digits 1 to 4 represent the four lexical tone categories in Mandarin Chinese, while “0” indicates the neutral tone.

To ensure the validity of the stimuli, we extracted the acoustic realization of the original full sentences using ProsodyPro (Y. Xu, 2013). We focused on the F0 realization of the stem-final T2 and T4, as well as the sentence-final neutral tone. The acoustic results (see Figure 1) showed F0 realizations of tone and intonation comparable to those reported in previous studies (Yuan, 2011; Zhang et al., 2022). Specifically, tones in question intonation were realized in a higher register than those in statement intonation, regardless of whether they were lexical tones or neutral tones and regardless of whether they were rising or falling. For the stem-final lexical tone in question intonation, T2 was realized as a steeper rise with an expanded range, while T4 was flattened with a reduced range.

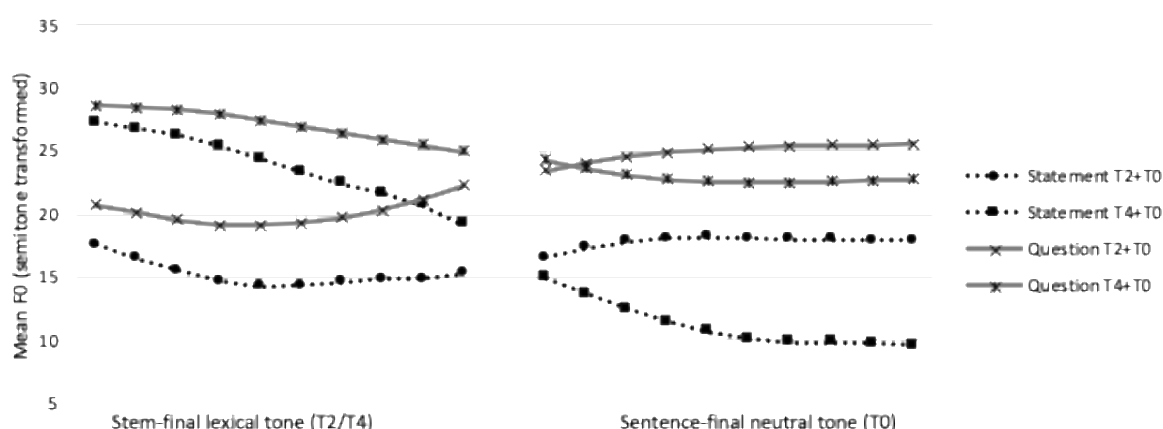


Figure 1. F0 contours of T2 and T4 in different stem-final positions within statements (dotted lines) and questions (solid lines). F0 contours are represented by time-normalized 10 equally spaced sample points. To normalize inter-speaker variations, all original F0 values in Hertz were converted to semitones with a reference of 64 Hz.

The recording yielded four statements and four questions from each speaker, which served as the Original stimuli. Two additional stimulus types were constructed using Praat. For the Truncated stimuli, the sentence-final *le* was excised from each original sentence, resulting in four statement stems and four question stems per speaker. For the Switched stimuli, the stem of a statement was adjoined with the final *le* from the corresponding question, and vice versa. In total, 48 experimental stimuli were constructed across the three conditions (3 stimulus types  $\times$  4 sentences  $\times$  2 intonation types  $\times$  2 speakers), with the 24 stimuli whose stem-final tones were T2 and T4 serving as testing stimuli while the other 24 stimuli whose stem-final tones were T1 and T3 served as fillers. Four additional yes-no intonational questions and their corresponding statements were used as practice stimuli to familiarize participants. The practice sentences were identical to the test SVO sentences except that they did not contain the SFP *le* (see Table A2 in the Appendix).

### 2.3 Procedures

Participants performed an intonation identification task programmed in PsychoPy (version 2021.3.4). The experiment consisted of one practice session followed by three main test sessions, with the progression between sessions self-paced. The practice session was designed to familiarize participants with the task procedure. The three main sessions each presented a distinct stimulus type, administered in a fixed order: the Truncated stimuli, the Switched stimuli, and the Original stimuli. This ordering was motivated by two reasons. First, presenting the full, naturally produced stimuli (Original) last prevents listeners from forming a memory-based standard against which the modified stimuli (Truncated and Switched) could be evaluated. Second, the sequence follows a progression from the most informationally reduced to the most informationally complete stimulus type. The Truncated stimuli present the sentence stem alone, without any sentence-final syllable; the Switched stimuli present both the stem and the final *le*, but drawn from different intonational contexts; and the Original stimuli present the full sentences. This progression allows listeners to gradually adapt to the task demands while minimizing carryover effects from more naturalistic to less naturalistic stimulus types. Within each session, 32 trials (4 sentences  $\times$  2 intonation types  $\times$  2 speakers  $\times$  2 repetitions) were presented in random order.

A two-alternative forced-choice paradigm was employed. Auditory stimuli were presented over headphones at a self-adjusted, comfortable listening level. Each trial began with a fixation cross (+) displayed for 500 ms, followed by the auditory stimulus. A brief on-screen instruction then prompted participants to judge whether the utterance was a statement or a question. Participants pressed the “1” key to indicate a question and the “0” key to indicate a statement. No response time limit was imposed, and the next trial began only after a response was made. No feedback was provided. All participants were tested individually in a quiet environment. The entire experiment (including the practice session) lasted approximately 10 minutes.

### 2.4 Data Analysis

We coded participant responses for each trial, assigning a value of 1 when the response was consistent with the intonation of the original sentence stem. The primary objectives of the

current analysis were to examine how native speakers use prosodic cues carried by the sentence stem and the final neutral-tone syllable in intonation identification and to investigate whether the rising or falling lexical tone at the sentence-stem-final position affects intonation identification. To avoid the interpretational complexity of a fully crossed three-factor design, the entire dataset was divided into two independent subsets based on the original sentence stem intonation: the statement subset and the question subset. For each subset, a binary generalized linear mixed-effects model (GLMM) was fitted with response-stem consistency (Consistency: 1=consistent, 0=inconsistent) as the dependent variable, as a function of two predictor variables: stimulus type (Type: original, truncated, switched) and stem-final lexical tone (Tone: T2, T4). Models included random intercepts for participants and by-participant random slopes for stimulus type. For significant interactions, post-hoc pairwise comparisons were conducted, with *p*-values adjusted using the *Tukey* method. All statistical analyses and post-hoc comparisons were conducted using the *lme4* package (Bates et al., 2015) in R version 4.4.1 (R Core Team, 2024), with the statistical significance threshold set at  $p < .05$ .

### 3. Results

All participants successfully completed the intonation identification task, and all recorded responses were included in the analysis. Figure 2 illustrates the mean rate of response-stem consistency across the different experimental conditions. Panel A displays the mean consistency rates in the statement context across three stimulus types, i.e., all of which share the statement stem, while Panel B presents the mean rates in the question context across the corresponding three stimulus types, i.e., all of which share the question stem. Overall, a consistent response pattern was observed. Specifically, listeners were able to accurately identify the intonation in both the original full sentence stimuli and the final-*le* truncated stimuli; however, in the switched condition, participants' perceptual judgments were almost inconsistent with the intonation of the sentence stem. Table 2 summarizes the results of the GLMMs for the statement-stem and question-stem subsets. In both models, the baseline was T2 in the original condition. Detailed results of the statistical analyses are reported and interpreted in the following subsections.

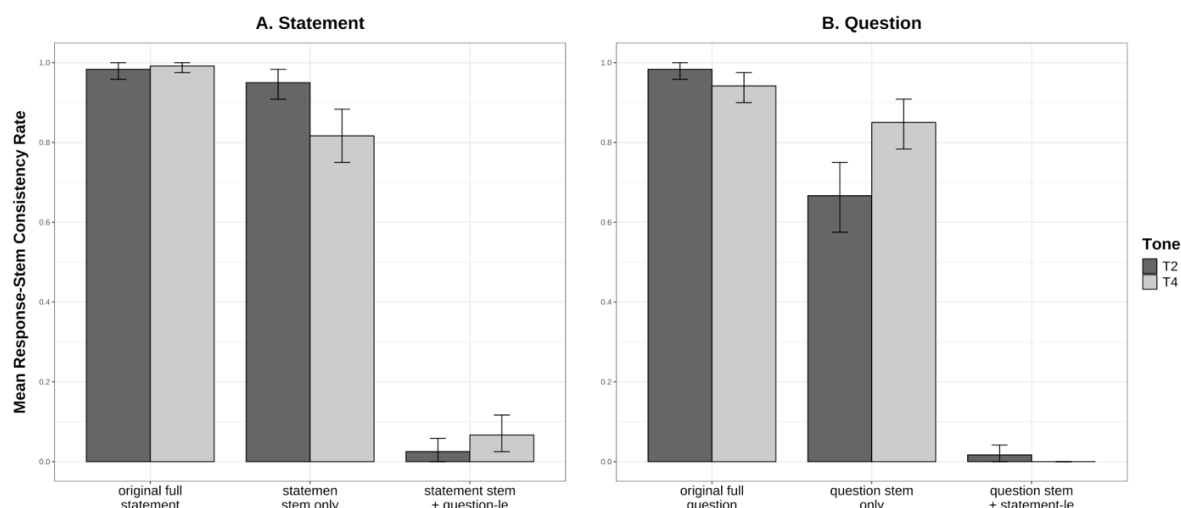


Figure 2. Mean consistency rates for statements (A) and questions (B) in the three stimulus types. Original stimulus type: original full statement and original full question; truncated stimulus type: statement stem only and question stem only; switched stimulus type: statement stem + question-le and question stem + statement-le.

Table 2. Summary of GLMMs for statement intonation and question intonation.

|               | Statement |       |          |          | Question |          |          |          |
|---------------|-----------|-------|----------|----------|----------|----------|----------|----------|
|               | Estimate  | SE    | <i>z</i> | <i>p</i> | Estimate | SE       | <i>z</i> | <i>p</i> |
| (Intercept)   | 8.801     | 2.557 | 3.44     | <.001    | 5.186    | 0.855    | 6.07     | <.001    |
| Truncated     | -5.736    | 2.555 | -2.25    | 0.025    | -4.204   | 0.834    | -5.04    | <.001    |
| Switch        | -12.797   | 2.62  | -4.88    | <.001    | -12.721  | 2.477    | -5.14    | <.001    |
| T4            | 0.891     | 1.451 | 0.61     | 0.539    | -1.424   | 0.889    | -1.6     | 0.109    |
| Truncated: T4 | -2.376    | 1.533 | -1.55    | 0.121    | 2.778    | 0.962    | 2.89     | 0.004    |
| Switch: T4    | 0.161     | 1.64  | 0.1      | 0.922    | -15.707  | 4058.115 | -0.01    | 0.997    |

*Note.* Model for response-stem consistency in the statement/question contexts: Consistency ~ Type \* Tone + (1 + Type | Participant). Statement model fit: marginal  $R^2 = 0.717$ ; conditional  $R^2 = 0.914$ . Question model fit: marginal  $R^2 = 0.909$ ; conditional  $R^2 = 0.971$ . The baseline was T2 in the original condition.

### 3.1 Results of Statement Intonation

The GLMM for the statement-stem subset revealed a highly significant main effect of Type,  $\chi^2(2)=165.14$ ,  $p<0.001$ , and a significant interaction between Type and Tone,  $\chi^2(2)=8.95$ ,  $p=0.011$ . The main effect of Tone was not significant,  $\chi^2(1)=2.37$ ,  $p=0.124$ .

Post-hoc pairwise comparisons were conducted to unpack the interaction (see Table A3 in the Appendix). Notably, the switched condition resulted in a dramatic reversal of perceptual judgments. Compared to the original condition, adjoining a question-final *le* onto the statement stem significantly decreased response-stem consistency for both the T2 and the T4 conditions ( $ps<0.001$ ). This indicates that participants overwhelmingly perceived the conflicting stimuli as questions, demonstrating that the intonational cue carried by the final neutral-tone syllable *le* is more likely to override that carried by the sentence stem.

Furthermore, the absence of the sentence-final *le* in the truncated condition interfered with intonation recognition, but this effect was modulated by the stem-final lexical tone. While removing the SFP led to a significant drop in consistency in the T4 condition ( $p=0.007$ ), the decrease was not significant in the T2 condition ( $p=0.064$ ). Moreover, in the truncated condition, consistency was significantly higher when the stem-final tone was T2 compared to T4 ( $p=0.003$ ), suggesting that statement intonation is more perceptually influenced by the loss of the sentence-final neutral-tone syllable when the stem-final full-tone syllable carries a falling lexical tone.

### 3.2 Results of Question Intonation

The GLMM for the question stem subset demonstrated significant main effects of Type,  $\chi^2(2)=33.16$ ,  $p<0.001$ , and Tone,  $\chi^2(1)=8.06$ ,  $p=0.005$ , as well as a significant two-way interaction between them,  $\chi^2(2)=8.35$ ,  $p=0.015$ .

Consistent with the patterns observed in the model for the statement stem subset, the switched condition triggered a reversal in intonation judgment. When a statement-final *le* was adjoined onto a question stem, response-stem consistency decreased for both the T2 and the T4 conditions. Notably, for T4-final question stems in the switched condition, standard errors became extremely large, indicating a state of complete statistical separation in which participants selected mismatched responses with absolute uniformity. That is, all participants perceived the conflicting stimuli as statements. This ceiling-level perceptual shift further supports the dominance of the statement-final *le* over the question stem.

The truncated condition revealed a mirror-image asymmetry to that found in the statement subset. While removing the question-final *le* caused a significant decrease in consistency for both the T2 condition ( $p<0.001$ ) and the T4 condition ( $p=0.022$ ), sentence stems ending with T4 syllables led to significantly higher response-stem consistency than those ending in T2 syllables ( $p<0.001$ ). This interaction implies that when the question-final neutral-tone syllable is removed, intonation judgment is highly influenced by the stem-final lexically rising tone.

## 4. Discussion

The present study aimed to determine the relative perceptual weights of the sentence stem

versus the sentence-final neutral-tone syllable in Mandarin intonation perception, and to investigate how the rising and falling lexical tones interact with intonation. By systematically removing and cross-splicing the sentence-final neutral-tone particle *le* across statement and question contexts, the results revealed a clear perceptual hierarchy, with a T2-T4 asymmetry in intonation judgment.

The current study found a dominant role of the final neutral-tone syllable over the sentence stem, as evidenced by the switched condition. When a statement stem was adjoined with a question-final *le* and vice versa, native listeners' perceptual judgments were highly reversed. They overwhelmingly categorized the stimuli based on the intonational information carried by the sentence-final *le*, ignoring the acoustic cues carried in the pre-final sentence stem. This perceptual override was so powerful that in the case of T4-final question stems appended with a statement-final *le*, it led to a complete statistical separation, i.e., all responses aligned with the statement-final *le*. While the sentence stem may exhibit global pitch modifications, these pre-final cues are not as perceptually efficient as the localized intonational cues on the sentence-final neutral-tone syllable. These findings provide solid evidence for the claim that the sentence-final syllable acts as the primary locus for intonational cues in Mandarin (Lin, 2006, 2012; B. Xu, 2013), and further demonstrate a dominant role of the final neutral-tone syllable (Jiang, 2010).

While the sentence-final neutral-tone syllable determines the ultimate intonational perception, the truncated condition revealed a highly systematic interaction between the stem-final lexical tone and intonation type. Specifically, in statements, removing the final *le* influenced the perception of T4-final stems significantly more than that of T2-final stems; in questions, however, the omission of the final *le* decreased question responses more for T2-final stems. This asymmetrical interference is consistent with previous studies (Yuan, 2011) and can be attributed to the dual function of F0 information in tone languages (Liang & Van Heuven, 2007; Torreira et al., 2014). Specifically, Mandarin T2 is a lexically rising tone, and question intonation is typically characterized by a final rise cross-linguistically. When a question ends in a T2 syllable, the F0 trajectories of this lexical tone and the question intonation share the same upward direction and high F0 target, creating perceptual ambiguity. As a result, listeners cannot reliably decode the final rise as a lexical rising tone or as question intonation. A parallel ambiguity occurs between T4 and statement. That is, both T4 and statement intonation have falling F0 trajectories, making it difficult for listeners to confirm the intonation type. Conversely, when stem-final lexical tone and stem intonation have conflicting trajectories, perceptual judgments were overall highly consistent with the stem intonation type. For example, final T2 and T4 usually undergo F0 modification caused by question intonation, with the former being raised and steeper while the latter being flattened. The intonation-induced modifications are highly marked and salient, and they do not cause tone-intonation ambiguity. Thus, even when the sentence-final toneless syllable is removed, listeners are able to identify the intonational cues carried by the stem more accurately.

These findings bring significant theoretical implications for Mandarin prosody, particularly regarding the role of syntactically independent neutral-tone syllables. The current data strongly suggest that although the phonetic realization of neutral-tone syllables such as the

SFP *le* is dependent on the preceding full-tone syllables, they are optimal carriers for intonational cues. In a tonal language like Mandarin, F0 information encodes both lexical tone and intonation, often leading to perceptual ambiguity. The neutral tone, on the other hand, offers a position for intonation realization. The T2-T4 asymmetry in intonation judgment in the truncated condition and the extremely low consistency in the switched condition both suggest that the final syntactically independent neutral-tone syllable is a better carrier of intonational cues and maximizes communicative efficiency.

## 5. Conclusion

The present study investigated the perception of intonation on neutral tone in Mandarin Chinese. It addressed two questions: whether the sentence-final neutral-tone syllable or the pre-final sentence stem serves as the primary perceptual anchor for intonation, and how the lexical tones of stem-final syllables interact with intonational cues. We hypothesized that the final neutral-tone syllable would perceptually dominate the sentence stem, and that this reliance would be modulated by the congruence or conflict between the stem-final lexical tone (T2 and T4) and the intonation type (statement and question).

Our behavioral findings strongly supported these hypotheses, revealing two primary insights. First, the neutral-tone SFP *le* exhibited perceptual dominance over the sentence stem, as substituting the particle completely reversed listeners' intonation judgments. Second, when the SFP *le* was removed, intonation judgments were modulated by the stem-final lexical tone, exhibiting a mirror-image asymmetry. Specifically, intonation identification was severely influenced when the F0 trajectory of the lexical tone was congruent with that of the statement or question intonation. These findings highlight the role of sentence-final syntactically independent neutral-tone syllables. By locating statement and question intonational cues on the final neutral-tone syllable, the prosodic system avoids ambiguity between word-level lexical tone and sentence-level intonation, thereby maximizing perceptual clarity.

However, several limitations must be acknowledged. First, the three stimulus types were presented in a fixed order, which introduces a potential confound. That is, differences in consistency rates across stimulus types may partly reflect practice and order effects rather than purely acoustic-perceptual factors. Second, the current study focused exclusively on perceptual judgments without corresponding acoustic production data, precluding a direct comparison of how intonation-induced F0 modifications in production map onto perceptual categories. Third, the stimuli were restricted to two lexical tones (T2 and T4) and a single sentence-final particle (*le*), which may not fully represent the diversity of the Mandarin prosodic system. Addressing these limitations will require several lines of future research. To distinguish stimulus-type effects from order effects, studies could employ a Latin-square counterbalancing scheme or a fully randomized design. To bridge the production-perception gap, future work should collect parallel acoustic data to examine intonation-induced F0 modifications in the stems and relate them to listeners' categorization patterns. To enhance generalizability, the experimental scope should be expanded to include all four lexical tones and other syntactically independent neutral-tone syllables. Finally, online measures such as eye-tracking or neurophysiological methods could provide deeper, multidimensional insights

into the temporal dynamics and neural mechanisms underlying tone-intonation interaction in Mandarin Chinese.

## Acknowledgments

This study was supported by Guangdong Planning Office of Philosophy and Social Science (GD23XWY13). We are deeply grateful to all participants who took part in our experiment.

## References

- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1-48. <https://doi.org/10.18637/jss.v067.i01>
- Boersma, P., & Weenink, D. (2023). *Praat: Doing phonetics by computer [Computer program]*. Retrieved from <http://www.praat.org/>
- Bolinger, D. (1982). Intonation and its parts. *Language*, 58(3), 505-513. <https://doi.org/10.2307/413847>
- Cao, J. (1992). On neutral-tone syllables in Mandarin Chinese. *Canadian Acoustics*, 20(3), 49-50.
- Chao, Y.-R. (1968). *A Grammar of Spoken Chinese*. University of Chicago Press.
- Chen, Y., & Xu, Y. (2006). Production of weak elements in speech – Evidence from F0 patterns of neutral tone in standard Chinese. *Phonetica*, 63(1), 47-75. <https://doi.org/10.1159/000091406>
- Cruttenden, A. (1981). Falls and rises: Meanings and universals. *Journal of Linguistics*, 17(1), 77-91. <https://doi.org/10.1017/S0022226700006782>
- Gussenhoven, C. (2002). Intonation and interpretation: Phonetics and phonology. *Speech Prosody*, 2002, 47-57. <https://doi.org/10.21437/SpeechProsody.2002-7>
- Gussenhoven, C. (2004). *The Phonology of Tone and Intonation*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511616983>
- Jiang, H. (2010). *Hanyu yudiao wenti de shiyan yanjiu [Experimental study on Mandarin intonation]*. Capital Normal University Press.
- Ladd, D. R. (2008). *Intonational Phonology* (2nd ed.). Cambridge University Press. <https://doi.org/10.1017/CBO9780511808814>
- Liang, J., & Van Heuven, V. J. (2007). Chinese tone and intonation perceived by L1 and L2 listeners. In C. Gussenhoven, & T. Riad (Eds.), *Tones and Tunes: Experimental Studies in*

*Word and Sentence Prosody* (Vol. 2, pp. 27-62). Mouton de Gruyter.  
<https://doi.org/10.1515/9783110207576.1.27>

Lin, M. (2006). Yiwen he chenshu yuqi bianjiediao [Interrogative vs. declarative intonation and boundary tone in Standard Chinese]. *Zhongguo Yuwen [Studies of the Chinese Language]*, 4, 364-376.

Lin, M. (2012). *Hanyu yudiao shiyan yanjiu [Experimental study on Mandarin intonation]*. China Social Sciences Press.

Liu, M., Chen, Y., & Schiller, N. O. (2016). Online processing of tone and intonation in Mandarin: Evidence from ERPs. *Neuropsychologia*, 91, 307-317.  
<https://doi.org/10.1016/j.neuropsychologia.2016.08.025>

Ohala, J. J. (1983). Cross-language use of pitch: An ethological view. *Phonetica*, 40(1), 1-18.

Ren, G.-Q., Tang, Y.-Y., Li, X.-Q., & Sui, X. (2013). Pre-Attentive Processing of Mandarin Tone and Intonation: Evidence from Event-Related Potentials. In F. Signorelli (Ed.), *Functional Brain Mapping and the Endeavor to Understand the Working Brain* (pp. 95-108). InTech. <https://doi.org/10.5772/56503>

Shi, X., & Li, H. (2025). Intonation-Induced F0 Realizations of Rising and Falling Lexical Tones in Mandarin Questions. *Open Journal of Modern Linguistics*, 15(4), 668-684.  
<https://doi.org/10.4236/ojml.2025.154039>

Tang, P., Yuen, I., Xu Rattanasone, N., Gao, L., & Demuth, K. (2018). Acquisition of weak syllables in tonal languages: Acoustic evidence from neutral tone in Mandarin Chinese. *Journal of Child Language*, 46(1), 24-50. <https://doi.org/10.1017/S0305000918000296>

Torreira, F., Roberts, S. G., & Hammarström, H. (2014). Functional trade-off between lexical tone and intonation: Typological evidence. In C. Gussenhoven, Y. Chen, & D. Dediu (Eds.), *Proceedings of the 4th International Symposium on Tonal Aspects of Language* (pp. 100-103). ISCA.

Xu, B. (2013). *Cross-linguistic perception of intonation by Cantonese and Mandarin listeners* [Master of Philosophy thesis, Chinese University of Hong Kong]. CUHK Digital Repository.

Xu, Y. (2013). ProsodyPro—A tool for large-scale systematic prosody analysis. *Proceedings of Tools and Resources for the Analysis of Speech Prosody (TRASP)*, 7-10.

Yuan, J. (2011). Perception of intonation in Mandarin Chinese. *The Journal of the Acoustical Society of America*, 130(6), 1063-1074. <https://doi.org/10.1121/1.2213572>

Zhang, Y., Schmidt, E., & Post, B. (2022). Perception of Intonation on Neutral Tone in Mandarin. *Frontiers in Communication*, 7, 849132.  
<https://doi.org/10.3389/fcomm.2022.849132>

## Appendix

**Table A1. Four Simple Conversations Embedding Test Sentences**

| No. | Sentence (Pinyin form)                      | Translation (statement version)         |
|-----|---------------------------------------------|-----------------------------------------|
| 1   | A <i>ni3 gang1gang1 shuo1 shen2me0?</i>     | ‘What did you say?’                     |
|     | B <i>zhang1hao4 zuo4 kao3rou4bing3 le0.</i> | ‘Zhanghao made grilled meat pie.’       |
|     | A <i>zhang1hao4 zuo4 kao3rou4bing3 le0?</i> | ‘Did Zhanghao make grilled meat pie?’   |
|     | B <i>shi4de0.</i>                           | ‘Yes.’                                  |
| 2   | A <i>ni3 gang1gang1 shuo1 shen2me0?</i>     | ‘What did you say?’                     |
|     | B <i>wang2bao3 chang2 dun4yu2tou2 le0.</i>  | ‘Wangbao tasted fish head stew.’        |
|     | A <i>wang2bao3 chang2 dun4yu2tou2 le0?</i>  | ‘Did Wangbao taste fish head stew?’     |
|     | B <i>shi4de0.</i>                           | ‘Yes.’                                  |
| 3   | A <i>ni3 gang1gang1 shuo1 shen2me0?</i>     | ‘What did you say?’                     |
|     | B <i>li3chao1 dian3 zha2ji1chi4 le0.</i>    | ‘Lichao ordered fried chicken wings.’   |
|     | A <i>li3chao1 dian3 zha2ji1chi4 le0?</i>    | ‘Did Lichao order fried chicken wings?’ |
|     | B <i>shi4de0.</i>                           | ‘Yes.’                                  |
| 4   | A <i>ni3 gang1gang1 shuo1 shen2me0?</i>     | ‘What did you say?’                     |
|     | B <i>zhao4hao2 chi1 shao1sun3gan1 le0.</i>  | ‘Zhaohao ate cooked bamboo shoots.’     |
|     | A <i>zhao4hao2 chi1 shao1sun3gan1 le0?</i>  | ‘Did Zhaohao eat cooked bamboo shoots?’ |
|     | B <i>shi4de0.</i>                           | ‘Yes.’                                  |

*Note.* The four digits 1 to 4 represent the four lexical tone categories in Mandarin Chinese, while “0” indicates the neutral tone.

**Table A2. Four Sentences (In the Statement Version and Question Version) in the Practice Session**

| No. | Sentence (Pinyin form)               | Translation (the statement version)   |
|-----|--------------------------------------|---------------------------------------|
| 1   | <i>zhang1hao4 zuo4 kao3rou4bing3</i> | ‘Zhanghao made grilled meat pie.’     |
| 2   | <i>wang2bao3 chang2 dun4yu2tou2</i>  | ‘Wangbao tasted fish head stew.’      |
| 3   | <i>li3chao1 dian3 zha2ji1chi4</i>    | ‘Lichao ordered fried chicken wings.’ |
| 4   | <i>zhao4hao2 chi1 shao1sun3gan1</i>  | ‘Zhao hao ate cooked bamboo shoots.’  |

**Table A3. Results of Post-hoc Pairwise Comparisons of Response-stem Consistency in Statement and Question Contexts**

|                        | Statement |       |        |        | Question |       |        |        |
|------------------------|-----------|-------|--------|--------|----------|-------|--------|--------|
|                        | Estimate  | SE    | z      | p      | Estimate | SE    | z      | p      |
| Origin: T2-T4          | -0.891    | 1.45  | -0.614 | 0.5392 | 1.42     | 0.889 | 1.603  | 0.109  |
| Truncated: T2-T4       | 1.485     | 0.494 | 3.006  | 0.0026 | -1.35    | 0.364 | -3.728 | 0.0002 |
| Switched: T2-T4        | -1.052    | 0.764 | -1.378 | 0.1682 | 16.31    | 2690  | 0.006  | 0.9952 |
| T2: Origin-Truncated   | 5.74      | 2.55  | 2.245  | 0.0638 | 4.2      | 0.834 | 5.041  | <.0001 |
| T2: Origin-Switched    | 12.8      | 2.62  | 4.885  | <.0001 | 12.72    | 2.48  | 5.136  | <.0001 |
| T2: Truncated-Switched | 7.06      | 0.787 | 8.97   | <.0001 | 8.51     | 2.29  | 3.713  | 0.0006 |
| T4: Origin-Truncated   | 8.11      | 2.67  | 3.036  | 0.0068 | 1.43     | 0.537 | 2.656  | 0.0216 |
| T4: Origin-Switched    | 12.64     | 2.7   | 4.673  | <.0001 | 28.43    | 4060  | 0.007  | 1.0000 |
| T4: Truncated-Switched | 4.52      | 0.479 | 9.437  | <.0001 | 27.00    | 4060  | 0.007  | 1.0000 |

*Note.* The Switched condition with T4-final question stems in the question subset exhibited extremely large standard errors ( $SE=4060$ ). This reflects a complete separation: all participants perceived the conflicting stimuli as statements. The Wald estimate,  $z$  value, and  $p$  value are statistically uninterpretable. The values are reported in the table for completeness but should not be relied upon for inference.

**Copyrights**

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>)