The Neutral Tone in Question Intonation in Mandarin

Fang Liu\(^1\), Yi Xu\(^2\)

\(^1\) Department of Linguistics, The University of Chicago, IL, USA
\(^2\) Department of Phonetics and Linguistics, University College London, UK

liufang@uchicago.edu, yi.xu@ucl.ac.uk

Abstract

This study investigates how the neutral tone, when preceded by different full tones under different focus conditions, behaves in question intonation in Mandarin. Results indicate that 1) the preceding full/neutral tone largely determines the local \(F_0\) trajectory of the neutral tone, but the latter also gradually converges over the course of several neutral tone syllables, 2) post-focus lowering, which is caused by the effect of focus, occurs in both neutral-tone-ending and High-tone-ending sentences, with the interrogative intonation in questions realized as an upward shift starting from the focused word, and 3) sentence-final neutral tone has a falling contour even in questions, thus contrasting with sentence-final High tone, which has a rising contour in questions.

Index Terms: neutral tone, focus, sentence type

1. Introduction

There are four full lexical tones and a neutral tone in Mandarin. On a five-point scale \([1]\), the High tone in citation form is transcribed as \(55\), the Rising tone \(35\), the Low tone \(214\), and the Falling tone \(51\). The pitch of the neutral tone is determined by its preceding tone: high after the Low tone, and relatively low after the other tones \([2,3]\). Nevertheless, in investigating the carryover and anticipatory effects of full tones on a sequence of 1-3 neutral tones, Chen & Xu \([4]\) concluded that the neutral tone has a static mid target, which is implemented with weak articulatory strength.

Although the basic characteristics of the four full tones in connected speech do not deviate much from their citation forms, there have been reports that intonation and sentence environment do modify their register and contour characteristics: 1) The overall pitch values of the four full tones in sentence-final position are higher in questions than in statements \([5-11]\). 2) The High tone falls moderately at the final position of a declarative sentence \([12,13]\), but rises in an interrogative sentence \([6,9,11]\). 3) The steepness of the Rising and Falling tone of the final syllable differs in questions and statements \([8]\), with the Rising tone steepened \([10]\) and widened \([7,9,11]\) and the Falling tone flattened \([10,11]\) and narrowed \([7]\) in pitch range in questions. 4) According to \([14]\), the sentence-final Low tone is realized as falling in statements but falling-rising in questions. However, in \([7]\) and \([11]\), the pitch realization of the final Low tone is falling-rising in both sentence types, except that the rising ending is higher in questions than in statements. The data in \([6]\) further reveal that only focused Low is realized as falling-rising statement-finally, and non-focused Low is falling in a statement-final position. On the other hand, the question-final Low tone is realized as falling-rising no matter whether it is focused or not.

The issue of tone-intonation interaction becomes even trickier when the neutral tone is involved. If indeed targetless as traditionally viewed \([2,3]\) or with a weakly implemented mid target \([4]\), the \(F_0\) of the neutral tone at the final position of a question should fully or at least largely reflect the interrogative intonation. Previous research, however, have produced mixed results. In Chao’s impressionistic account \([2]\), for a “Low + High + Rising + Neutral + Neutral + Neutral” sequence, \(\text{“nǐ \(21\)} shuō \(55\) shēn \(35\) me \(21\) lāi \(51\) zhe \(1/1\) [\text{What did you say?}]\), the last three neutral tones are realized with a low pitch if the question is newly-uttered, but with a high pitch (“nǐ \(21\) shūō \(55\) shēn \(35\) / me \(21\) lāi \(51\) zhe \(5/5\)?) if the same question is asked again. Qi \([15]\), however, stated that the neutral-tone-question particle “\(ma\)” has a high pitch /5/ no matter what the preceding full tone is. Furthermore, with a sequence of three neutral tones plus “\(ma\)” at the end of a question, their pitch can be transcribed as \(444\) /5/ regardless of what the preceding full tone is. Yet, upon observing that the question-final neutral-tone particle “\(ma\)” has a falling contour after the High and Falling tone, a rising contour after the Low tone, and a falling/level/rising contour after the Rising tone, Shen \([9:48]\) claimed that the tonal value of the sentence-final interrogative particles is “the algebraic sum of the \(F_0\) value of the preceding tone and the sentence intonation”, and that they “always end on a high key”. In the Pan-Mandarin ToBI system \([16]\), the sentence-final particle “\(ma\)” can carry two possible boundary tones: H% in questions and L% in statements. However, Lee \([17:122]\) proposed that “the \(ma\)-particle does not stand as an independent prosodic unit that could bear the terminal rise” in marked questions.

In summary, previous research on Mandarin has not reached a consensus on the pitch pattern of the neutral tone in interrogative sentences, and the picture presented is far from complete. There was no explicit indication of how the preceding tone and the sentence intonation interact to generate the \(F_0\) contour of a question-final neutral tone in the algebraic sum hypothesis in \([9]\). The limited examples in \([16]\) and \([17]\) also failed to present a full picture of the influence of different preceding tones on the neutral tone in different intonations. Chen & Xu \([4]\) have neither examined the neutral tone at the sentence-final position nor in questions. Therefore, this study attempts to address the following unresolved issues regarding the neutral tone in question intonation in Mandarin: (1) Which of the two has a stronger influence on the \(F_0\) trajectory of the neutral tone: the preceding tone or sentence type? (2) Does the neutral tone have different targets in different sentence types? (3) What is the effect of focus on the neutral tone syllables in statements and questions? (4) Is the neutral tone more effective than a full tone in manifesting the statement/question distinction?

2. Method

2.1. Materials

Tā māi māma/yéye/nǎinai/méimei men de le ma
[He bought mothers’/grandpas’/grandmas’/sisters’]

Tā māi māma/yéye/nǎinai/méimei men de māomā
[He bought mothers’/grandpas’/grandmas’/sisters’ kitten]
The above two sentence frames are employed in the experiment, onto which the following three conditions are imposed to create 32 combinations.

- **Focus: nonadjacent vs. adjacent.** The focus is either (on māi) separated from the sequence of 5 neutral tones (or 3 neutral tones + 2 High tones) by a full tone, or (on māyié/nái/méi) immediately adjacent to them.
- **Preceding tone: High, Rising, Low, or Falling.** The sequence of 5 neutral tones (or 3 neutral tones + 2 High tones) is preceded by each of the four full tones.
- **Sentence type: statement vs. question.** Each utterance with the same components was produced with two alternate sentence types.

Each utterance was repeated five times by each subject, resulting in a total of 1280 sentences. The intended focus and sentence type were elicited by proper leading sentences.

### 2.2. Subjects

Eight native speakers of Mandarin, 4 females and 4 males, served as subjects. They were either students at Yale University or residents in New Haven, Connecticut, who were born and raised in the city of Beijing where Mandarin is the vernacular. They had no self-reported speech and hearing disorders and their ages ranged from 23 to 34.

### 2.3. Procedure

Recording was done in a sound-treated booth at Haskins Laboratories, New Haven, Connecticut. A JavaScript program displayed the target sentences (and the leading sentences) one at a time on a computer screen in random order. The recorded utterances were digitized at 44.1 kHz sampling rate and 16-bit amplitude resolution, and were later re-sampled at 22.05 kHz. For visual inspection and graphic analysis, a custom-written script for the Praat program [18] computed time-normalized F0 contours of the sentences by getting the same number of evenly spaced F0 points from each syllable.

### 3. Results

To see how sentence type is affected by lexical tone and focus, time-normalized F0 contours of the statements and questions with focus on māi and māyié/nái/méi are displayed in Fig. 1 and 2, respectively, each averaged across 40 repetitions by eight speakers.

As can be seen, regardless of the tone of the third (māyié/nái/méi) and last two syllables (neutral vs. High), in each graph, questions differ from statements with an increasingly higher F0 starting from the focused syllable (māi in Fig. 1 and māyié/nái/méi in Fig. 2) till the end of the sentence. Furthermore, this difference is larger in the final portion of the sentence when the last two tones are High than when they are neutral. Regarding the effect of focus, comparing Fig. 1a-d with Fig. 2a-d, we can see that focused māi not only lowers its own F0 but also raises the F0 of the preceding syllable tā, whereas focused māyié/nái/méi expands its own pitch range as well as that of the following neutral-tone syllable yē/nái/méi.

In Fig. 1c and 2c, due to the tone sandhi rule (Low > Rising > Low) in Mandarin [3], the second syllable māi (Low) changes into māi (Rising) before the third syllable nái (Low). Because focus is on māi in Fig. 1c but on nái in Fig. 2c, the rising contour of ná is more fully realized in the former than in the latter. Furthermore, the F0 trajectories of the rest of the syllables differ dramatically in the two figures because of the different focus locations. In Fig. 1c, F0 of the neutral-tone syllable ná remains low until the post-focus Low-tone syllable ná, but that of the next neutral-tone syllable men increases slightly, which is in turn followed by a slow decrease in the rest of the neutral-tone syllables de le ma. In Fig. 2c, however, following the focused Low-tone syllable ná, F0 of the next two neutral-tone syllables ná men rises sharply for the most part, but starts to fall during de and continues to fall throughout le ma. This phenomenon is described as "post-L F0 rise" in [4]. The current finding suggests that this effect is quite weak when the Low tone itself is post-focus.

To see how the neutral tone sequences (and the 3-neutral plus 2-High sequences) behave after each of the four full tones under different focus and sentence type conditions, F0 contours of the statements and questions alternating across māmiá/yē/e/nái/ná/méi/mei/me/mei/me/mei on the third and fourth syllable with the neutral-tone and High-tone ending are displayed in Fig. 3 and 4, respectively, each averaged across 40 repetitions by eight speakers.

---

Fig. 1: F0 contours of statements/questions with focus on “māi”. In the legend, “S/Neutral” refers to a statement ending with 5 neutral tones, “Q/High” a question with 2 High tones, and so on. The title “[ma3] ma1ma” indicates that focus is on “māi” and the 3rd and 4th syllables of the sentence are “māma”.

Fig. 2: F0 contours of statements/questions with focus on “māyié/nái/méi”. The title “ma3 [ma1ma]” indicates that focus is on “māma”, the 3rd and 4th syllables of the sentence.
What is observed above for the neutral tone sequence also applies to the 3 neutral tones plus 2 High tones sequence in Fig. 4, except that the High-tone ending acquires a slight rising contour in questions, as opposed to its level pattern in statements.

A set of repeated measures ANOVAs were conducted using R [19] to see if the observations made on Fig. 1-4 are statistically significant. In each of the ANOVA models, mean $F_0$ (in semitone) of each of the eight syllables in the sentences is treated as the dependent variable, sex as the between-subjects factor, and focus, sentence type, final tone, and preceding tone as the within-subjects factors. Results of the main effects from these models are shown in Table 1.

### Table 1. Results of the main effects from repeated measures ANOVAs of mean $F_0$ of each syllable on sex (Female vs. Male), focus (Nonadjacent: on “māi” vs. Adjacent: on “māi”), sentence type (Question vs. Statement), final tone (High vs. Neutral), and preceding tone (High, Rising, Low, or Falling). The $p$-values less than 0.05 are bolded.

<table>
<thead>
<tr>
<th></th>
<th>sex</th>
<th>focus</th>
<th>sentence type</th>
<th>final tone</th>
<th>preceding tone</th>
</tr>
</thead>
<tbody>
<tr>
<td>tā</td>
<td>F</td>
<td>M</td>
<td>A</td>
<td>N</td>
<td>Q</td>
</tr>
<tr>
<td>p</td>
<td>0.0079</td>
<td>0.0320</td>
<td>0.1596</td>
<td>0.1008</td>
<td>0.1738</td>
</tr>
<tr>
<td>māi</td>
<td>F</td>
<td>M</td>
<td>A</td>
<td>N</td>
<td>Q</td>
</tr>
<tr>
<td>p</td>
<td>0.0058</td>
<td>0.9996</td>
<td>0.0297</td>
<td>0.0094</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>māi</td>
<td>F</td>
<td>M</td>
<td>A</td>
<td>N</td>
<td>Q</td>
</tr>
<tr>
<td>p</td>
<td>0.0104</td>
<td>0.0075</td>
<td>0.0023</td>
<td>0.0605</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>māi</td>
<td>F</td>
<td>M</td>
<td>A</td>
<td>N</td>
<td>Q</td>
</tr>
<tr>
<td>p</td>
<td>0.0039</td>
<td>0.0002</td>
<td>0.0004</td>
<td>0.2200</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>de</td>
<td>F</td>
<td>M</td>
<td>A</td>
<td>N</td>
<td>Q</td>
</tr>
<tr>
<td>p</td>
<td>0.0019</td>
<td>0.0003</td>
<td>0.0001</td>
<td>0.2274</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>le</td>
<td>F</td>
<td>M</td>
<td>A</td>
<td>N</td>
<td>Q</td>
</tr>
<tr>
<td>p</td>
<td>0.0012</td>
<td>0.0141</td>
<td>&lt; 0.0001</td>
<td>0.0365</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>māi</td>
<td>F</td>
<td>M</td>
<td>A</td>
<td>N</td>
<td>Q</td>
</tr>
<tr>
<td>p</td>
<td>0.0008</td>
<td>0.5456</td>
<td>&lt; 0.0001</td>
<td>&lt; 0.0009</td>
<td>&lt; 0.0001</td>
</tr>
</tbody>
</table>

The main effects shown in Table 1 can be summarized as follows. 1) Female speakers generally have higher $F_0$ than male speakers. 2) The effect of focus varies with its location: (i) When it is on māi, which is nonadjacent to the following neutral tone sequence, focus has an anticipatory raising effect on the mean $F_0$ of the preceding syllable tā. (ii) When focus is on māi, which is adjacent to the following neutral tone sequence, it raises the mean $F_0$ of both the focused and the post-focus syllables. 3) Starting from the focused item, the $F_0$ of questions becomes higher than that of statements. The difference in $F_0$ height between the two is increasingly larger as the sentences approach the end. This is demonstrated by the more and more significant effect of sentence type on the mean $F_0$ of the eight consecutive syllables in the sentences. 4) There is no obvious anticipatory effect of the tone of the final two syllables (High vs. neutral) on the mean $F_0$ of the preceding syllables. Rather, the effect is strictly local: sentences ending with High tones have significantly higher final mean $F_0$ than those ending with neutral tones.
5) According to the tone sandhi rule \((L + L > R + L)\) in Mandarin, mái \((L)\) becomes mái \((R)\) before náï \((L)\), which is why the mean \(F_0\) of mái is the highest before the Low tone (when compared to the other three). Following mái, the mean \(F_0\) of the four full tones shows the order of méi > náï > má > yé. This is due to the articularity constraint that determines how fast \(F_0\) can change even with full effort [20]. Following the four full tones, there are four corresponding neutral tone syllables whose \(F_0\) trajectory seems to be a continuation of its preceding full tone, thus having mean \(F_0\) in the order of mái > méi > ye > náï. Following the ending point of máï/yé/ye/náï, the \(F_0\) contour of the neutral tone syllable men is mostly falling after máï/yé/ye/náï, but mostly rising after náï (especially when náï is focused). This makes its mean \(F_0\) the highest after yé, but much lower after máï/náï/náï/náï. The \(F_0\) contour of de is falling no matter what the preceding full tone is, but the “post-L \(F_0\) rise” of the focused náï on nai men raises the pitch range of their following de so that this de has a higher mean \(F_0\) than those preceded by the other three full tones. Although most cases of the last two syllables le māi/māmōnì converge to a falling/level contour in statements, only those preceded by mái men de/méiméi men de converge in questions. In addition, le mā/māmōnì are in a relatively higher pitch range when following focused/unfocused yé ye men de and focused náï men de.

4. Discussion and Conclusions

Conventionally considered to be “targetless”, the neutral tone at the question-final position has been assumed to be fully free to carry the rising intonation and therefore should be high in pitch [15]. However, the present study suggests that the neutral tone is not necessarily any better than a full tone in manifesting question intonation, because it does not exhibit a high pitch in a question, but rather behaves differently according to the tone and focus conditions of the preceding syllables. Furthermore, post-focus lowering occurs in the neutral-tone-ending sentences in both statements and questions, just as in the High-tone-ending sentences.

In conclusion, this study shows that 1) the \(F_0\) trajectory of the neutral tone is mainly determined by the preceding tone, which itself is heavily influenced by the focus and intonation conditions of the sentence; 2) when the \(F_0\) contours of the neutral tone in different conditions converge over time, they appear to be slightly falling in both statements and questions. Nonetheless, the fall is steeper in statements than in questions; 3) the effects of focus and sentence type are exerted on the neutral tone via global pitch controls. In this experiment, the neutral tone sequence occurs in the post-focus region and at the sentence boundary. Thus, the post-focus pitch range suppression due to focus and the nonlinear post-focus pitch increase due to question intonation [6] combine to form the pitch pattern of the neutral tone sequence (and the neutral tone 1 High tone sequence); 4) the \(F_0\) of the sentence-final High tones remains level in statements, but becomes rising in questions. This is in contrast to the \(F_0\) of the final neutral tones, which keeps falling not only in statements but also in questions. These patterns indicate that, in Mandarin, sentence type is an independent intonational function whose manifestation has to be achieved through the articulatory implementation of local tonal pitch target specified by the lexical tones, including the neutral tone.

5. References