Identifying Mandarin Tones with Intonation is Difficult for L2 Learners, Regardless of Their Experience

Wei William Zhou

The Ohio State University, USA
zhou.2874@osu.edu

Abstract

In Mandarin Chinese, fundamental frequency (f0) is the primary cue for the identification of both lexical tone and intonation. When tone and intonation coincide, the f0 parameters of intonation-overlaid tones deviate from those of citation tones. In this study, a tone identification experiment was conducted with both inexperienced and experienced second language (L2) learners to investigate how well L2 learners identify tones produced in two different intonation patterns (statement and question) and whether increased L2 experience facilitates the identification. The results showed that tones overlaid by question intonation were more difficult to identify than those not. Moreover, the experienced learners did not outperform the inexperienced learners, suggesting that identifying intonation-overlaid tones is difficult for learners, regardless of their L2 experience.

Index Terms: Mandarin Chinese, tone identification, intonation, L2 tone and intonation

1. Introduction

Standard Mandarin Chinese (Mandarin hereafter) has four lexical tones to distinguish word meaning, and f0 is the primary acoustic parameter for tone characterization [1]. In terms of f0 contour and height (on a scale of 1-5, 5 indicating the highest), Mandarin tones can be described as Tone 1 high-level (55), Tone 2 high-rising (35), Tone 3 low-dipping (214), and Tone 4 high-falling (51) [2]. Intonation, which conveys contextual information (e.g. statement, question, focus), also makes use of pitch variation in Mandarin [3]. While lexical tone is carried out by individual syllables and intonation is realized at the utterance level [4], the two coincide, raising the question of how tone and intonation function in Mandarin.

Chao [5, pp. 49–50] compared lexical tone and sentence intonation in Mandarin with "small ripples riding on large waves (though occasionally the ripples may be ‘larger’ than the waves)," and the result of the interaction is an "algebraic sum" of the two. This pioneering theory, though impressionistic, has motivated investigations of Mandarin intonation. Question intonation in Mandarin demonstrates both global and localized f0 effects. Proposals on the global aspect of question intonation include Gärding’s [6] grid model (rising grid for questions and falling grid for statements), X. Shen’s [7] two-layer model (an upper layer for questions and a lower layer for statements), J. Shen’s [8], [9] top line and base line model (gradually falling top line and rising/narrowing base line for questions, and sharply falling top line and falling/widening base line for statements), and Yuan’s [10], [11] "phrase curve" proposal (overall higher phrase curves for questions). On the other hand, proposals on localized f0 effects of intonation differ in where the primary acoustic cue lies. Except for a few studies arguing that question intonation begins with a higher tone register than does statement intonation, e.g. [7], [12], much experimental work has found that the relevant acoustic difference between the two intonation types occurs at the end of the utterance in Mandarin, e.g. [13]–[15]. Despite the disagreement on the location within the utterance that bears the most crucial acoustic information, there seems to be a consensus among linguists that the utterance-final tones are affected by intonation. With respect to how the utterance-final tones are affected by question intonation in Mandarin, there are different proposals. Chao [16] proposed the tone addition rule, in which question intonation raises and expands the final pitch of the citation tones, i.e. Tone 1 (T1) 55 → 56, T2 35 → 36, T3 214 → 216, and T4 51 → 513. On the other hand, Shen’s [7], [17] research supported the pitch neutralization rule, under which question intonation raises the f0 values of tones, but the basic shape and feature of the tones are preserved. Specifically, question intonation changes T1 from high-level into a rising slope and widens the pitch range of a rising tone but not that of a falling tone. In this process, no prolongation or addition is observed. In another series of studies, Yuan [10], [11] proposed a tone-dependent mechanism whereby question intonation flattens the falling slope of the falling T4 and steepens the rising slope of the rising T2; however, Yuan’s proposal did not address T1 and T3.

While studies on tone and intonation interaction propose different mechanisms by which question intonation affects tones, they concur that question intonation alters the f0 information of the utterance-final tones to a certain extent. An interesting question then arises as to how the presence of intonation can affect speakers’ perception of these tones. On this account, there has been evidence that native speakers of Mandarin can competently identify intonation-overlaid tones, reaching the “ceiling effect” [18]; however, it is not yet clear to what extent this can be applied to L2 speakers. Yang and Chan [19] investigated this question and discovered that native speakers and advanced English-speaking L2 learners of Mandarin identified tones in questions more accurately than did first- and second-year learners, and that tones in questions were harder to identify than tones in statements. However, several questions remain unanswered in Yang and Chan’s study. Due to including the native speaker control group in their analytical model, it is not explicit whether the better performance of the native/advanced L2 speaker group was primarily driven by the native speakers. The first- and second-year learners appeared to show no difference in identifying tones in questions and those in statements, potentially exposing their underdeveloped tone recognition ability; thus, it seems that intonation only adversely affects more advanced speakers in their identification of tones, begging the question of at what stage intonation begins to affect L2 speakers.

In summary, previous research on the Mandarin tone and intonation acquisition did not provide sufficient evidence to
address the manner of and extent to which intonation affects tone identification by L2 speakers. There is also no clear indication of when and how L2 speakers at various proficiency levels differ in their identification of intonation-overlaid tones. Therefore, this study aims to answer two questions: (1) How well can English-speaking L2 learners of Mandarin identify tones overlaid by question intonation? (2) Does the increase of L2 experience help with the identification?

2. Method

2.1. Stimuli

The stimuli were 96 syllables, including 12 syllable types (/pa/, /pia/, /ta/, /tia/, /mu/, /mua/, /ni/, /nia/) in combination with four lexical tones (T1, T2, T3, and T4) and two intonation types (statement and question). The syllables appeared in the sentence-final position of the following carrier sentence:

Zhangsan shuo ____/?!

‘Zhangsan says ____/?’

A trained, female native speaker of Mandarin was instructed to read the sentences with the appropriate statement and question intonation patterns. After the speech was obtained, the utterance-final syllables were extracted from their host sentence in Praat [20] in preparation for the experiment.

The stimuli were not normalized to ensure intonation authenticity. In the stimuli, there was a significant duration difference between tones in statements (M = 423 ms, SD = 90 ms) and tones in questions (M = 518 ms, SD = 97 ms), t(47) = 7.93, p < 0.001. The duration difference between the two intonation conditions holds true for all four tones: (1) T1s in statements (M = 421 ms, SD = 41 ms) and in questions (M = 471 ms, SD = 83 ms), t(11) = 3.14, p = 0.005; (2) T2s in statements (M = 459 ms, SD = 53 ms), and in questions (M = 502 ms, SD = 79 ms), t = 2.74, p = 0.01; (3) T3s in statements (M = 500 ms, SD = 58 ms) and in questions (M = 600 ms, SD = 75 ms), t = 5.97, p < 0.001; and (4) T4s in statements (M = 311 ms, SD = 70 ms) and in questions (M = 498 ms, SD = 104 ms), t = 8.62, p < 0.001. There was also an intensity difference between tones in statements (M = 68 dB, SD = 5 dB) and those in questions (M = 69 dB, SD = 5 dB), t(47) = 2.10, p = 0.04. However, in terms of individual tones, only T2s and T3s, but not T1s or T4s, exhibit an intensity difference between the two intonation conditions: (1) T2s in statements (M = 66 dB, SD = 2 dB) and in questions (M = 68 dB, SD = 2 dB), t(11) = 3.08, p = 0.01; T3s in statements (M = 63 dB, SD = 3 dB) and in questions (M = 64 dB, SD = 2 dB), t(11) = 2.33, p = 0.04; (3) T1s in statements (M = 72 dB, SD = 4 dB) and in questions (M = 74 dB, SD = 4 dB), p = n.s.; and (4) T4s in statements (M = 73 dB, SD = 3 dB) and in questions (M = 71 dB, SD = 4 dB), p = n.s. The longer duration and higher intensity of tones in questions are in line with the previous literature, e.g. [11].

2.2. Participants

The participants were 16 American, English-speaking L2 learners of Mandarin, including eight inexperienced learners (3M, 5F, aged 18-22) and eight experienced learners (4M, 4F, aged 24-39). All inexperienced learners had less than three years of L2 Mandarin experience and were students in a Mandarin language class, and all the experienced learners had over five years of L2 experience and were Mandarin language instructors or researchers at a major U.S. institution at the time of the experiment.

Two native Mandarin speakers completed the task and reached 100% accuracy. With this result along with the finding that native speakers reached the “ceiling” effect when identifying intonation-overlaid tones [18], no native control group was deemed necessary in this study.

2.3. Procedure

All the participants listened to the 96 stimulus items in a pseudo-randomized order with a three-second interstimulus interval. For each item, they chose the heard tone for the item from four options (T1, T2, T3, or T4). They only heard each item once. The participants completed the experiment online through a website hosted by a university server.

3. Results

3.1. Accuracy by intonation and L2 experience

In the design, the two predictor variables are intonation (tones in questions vs. in statements) and L2 experience (inexperienced vs. experienced). The outcome variable is response accuracy (correct vs. incorrect). Figure 1 shows the response accuracy by intonation and L2 experience. The data were fit into a logistic mixed-effects regression model in R [21] with the lme4 package [22]. The model predicted accuracy from intonation and L2 experience. The model was fit with the maximal random effects structure first [23] but resulted in a convergence error. After the random slopes were removed one at a time based on the random effect correlations and variances, the model with a random subject intercept and a random item intercept converged.

A log-likelihood comparison revealed a significant main effect of intonation type (χ²(1) = 14.68, p < 0.001), not of L2 experience. That is, participant identification performance for tones in questions is reliably poorer than that for tones in statements, but performance between the experienced learners and the inexperienced learners in the two intonation conditions did not differ. This suggests that for both experienced and inexperienced L2 learners, identifying tones in questions is more difficult than doing so for tones in statements.

Figure 2: Accuracy by intonation and L2 experience.
3.2. Accuracy by tone and intonation

There were discernible performance differences on the four lexical tones. Table 1 shows the identification accuracy by tone and intonation with data from the two learner groups aggregated.

<table>
<thead>
<tr>
<th>Tone</th>
<th>Statement</th>
<th>Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (high-level)</td>
<td>0.97</td>
<td>0.46</td>
</tr>
<tr>
<td>2 (high-rising)</td>
<td>0.61</td>
<td>0.74</td>
</tr>
<tr>
<td>3 (low-dipping)</td>
<td>0.97</td>
<td>0.97</td>
</tr>
<tr>
<td>4 (high-falling)</td>
<td>0.99</td>
<td>0.79</td>
</tr>
</tbody>
</table>

While participant performance is high and consistent for T3 in both intonation conditions, the performance for T1 and T4 shows the opposite pattern from T2. Specifically, for T1 and T4, accuracy is higher in statements than in questions, but for T2, accuracy is higher in questions than in statements. This means that while T1 and T4 in questions are harder for participants, T2 in questions is easier. This suggests that L2 identification of intonation-overlaid tones is tone-dependent, possibly based on whether their f0 encodings are congruent with the mechanisms of the question intonation [18], [24], [25]. When the target tone is a rising tone (T2), its tone features are in harmony with the question intonation, resulting in better identification; however, when the target tone is a falling tone (T4), its tone features conflict with the mechanisms of the question intonation, contributing to poor identification.

3.3. Error patterns by tone

Among the eight conditions of combined tone and intonation in Table 3, four conditions had an accuracy rate lower than 80%. They are: (i) T1 in questions, (ii) T2 in statements, (iii) T2 in questions, and (iv) T4 in questions. The errors in (i)-(iii) showed consistent patterns. In (i), T1s were mostly misidentified as T2s (91%), meaning that participants perceived the high-level tone as a rising tone. This pattern is in part consistent with the previous acoustic research wherein question intonation reportedly transforms T1 from level into a rising slope [17]. In both (ii) and (iii), T2s were predominantly misidentified as T3s (99% and 96%). This is in accordance with the well-established phenomenon of T2-T3 confusion by speakers of Mandarin, both native [26], [27] and non-native [28]. In the same respect, there seems to exist a global T3 bias in this study, which is that T2s were mostly misidentified as T3s but not contrariwise. One possible reason for this is the noticeable “creaky” phonation and long duration of T3s, which could have served as salient cues for the listeners. Condition (iv) did not show specific error patterns, where the target T4s were confused with all three other tones.

4. Discussion

In this study, both inexperienced and experienced L2 learners of Mandarin did well with tones in statements but struggled with tones in questions. This is understandable because in an L2 learning context, tones are taught and learned with a generic intonation, i.e. tones in statements, and the interaction between tone and intonation is rarely stressed. The inexperienced learners and the experienced learners did not exhibit any difference in their identification of tones in both statements and questions. This is different from the patterns observed in Yang and Chan [19] whereby inexperienced learners (first- and second-year learners) did not suffer from the influence of intonation when identifying intonation-overlaid tones. This discrepancy suggests that in a tone identification task of this nature, learner’s proficiency level is crucial for an intonation effect to be observed.

This study also found that when identifying intonation-overlaid tones, L2 speakers are subject to a tone-dependent mechanism. That is, L2 identification of intonation-overlaid tones can be hindered or facilitated by tone identity. While T4s in statements were identified less accurately than were T4s in question, T2s in questions were identified more accurately than were T2s in statements. However, this difference based on tone identity was not seen among native Mandarin speakers, whose identification of intonation-overlaid tones is consistent regardless of tone identity, as seen in [18] and in this current study (i.e. the two native speakers who completed the task with 100% accuracy). This points to the possibility that native speakers and L2 speakers attend to different aspects of the acoustic information in their identification of intonation-overlaid tones. The specific acoustic cues used by native and L2 speakers are unclear and go beyond the scope of this study. However, as previous research has suggested, English speakers tend to attach more importance to f0 height in tone identification while Mandarin speakers tend to place emphasis on both f0 height and direction [29]. The results in this study seem to suggest this tendency. Identification of intonation-overlaid tones by English-speaking L2 speakers of Mandarin appears to be less holistic and can be easily hindered or facilitated by tone identity.

Interestingly, native Mandarin speakers showed the tone-dependent effect in a sentence intonation type identification task, i.e. listening to an utterance and judging whether the utterance is a statement or a question [24], [25]. In such a task, sentences ending with a rising tone are harder for native speakers to identify as questions, possibly due to the fact that the f0 encodings of the rising tone coincide with the question intonation. Conversely, sentences ending with a falling tone are easier for speakers to perceive as questions, because the f0 encodings of the falling tone conflict with the question intonation. However, it is not clear whether L2 speakers will also demonstrate this tone dependence in a task of this kind, and whether more experienced speakers would have an advantage over less experienced speakers.

5. Conclusions

This study shows that the interaction of tone and question intonation in Mandarin affects L2 speakers’ identification. This effect is tone-dependent: question intonation negatively affects the identification of level tone (T1) and falling tone (T4) but facilitates that of rising tone (T2). The results may possibly be due to whether the f0 encodings of the tones are in harmony or in conflict with the question intonation. The experienced L2 speakers, who were instructors/researchers of Mandarin, did not perform better than the inexperienced speakers, suggesting that tone identification under the influence of question intonation is challenging for L2 speakers regardless of their experience.

6. Acknowledgements

I am grateful to Marjorie K.M. Chan, Mineharu Nakayama, Björn Köhnlein, and Nicole Nicholson for discussion and comments, to the Department of East Asian Languages and
Literatures at The Ohio State University for providing financial support for conference expenses, and to the audience at the 1st International Conference on Tone and Intonation for their comments and encouragement.

7. References


