The effect of pitch interplay on English-speaking learners of Mandarin

Shan Luo
University of Victoria

This study examines how native English speakers perceive and produce intonation and tone in Mandarin statements and unmarked questions. The results showed, as predicted, that English speakers had less difficulty perceiving intonation and sentence-final tone when the pitch movement of both was in the same direction than when it was not. On the production side, English speakers performed much better at producing tone than intonation. The intelligibility of their question intonation was especially compromised, likely due to their narrow pitch range. The English speakers also consistently produced more final rising pitch in Mandarin unmarked questions, suggesting phonological influence from their native language. Findings highlight the role of language experience in language learning, and the importance of pedagogical awareness.

Keywords: intonation, tone, unmarked questions, Mandarin, pitch interplay

1. Introduction

1.1 Background

Non-native (L2) use of intonation is a much-discussed topic in the L2 acquisition literature. Using an inappropriate intonation pattern gives rise to misunderstandings or negative impressions of L2 speakers (see Backman, 1979; Barlow, 1998; Broselow, Hurtig, & Ringen, 1987; Celce-Murcia, Brinton, & Goodwin, 1996; Chun, 2002; Cruz-Ferreira, 1987; Lepetit, 1989; Major, 2001; Mennen, 2004, 2007; McGory, 1997; Rasier & Hiligsmann, 2007; Ueyama, 2000; Ueyama & Jun, 1996; Vaissière, 2005). Studies have observed that the influence of native (L1) language experience is reflected both at the phonetic level in L2 production, shown as different pitch range originating from L1 (e.g., Barlow, 1998; Mennen, 2007; Ueyama, 2000; Willems, 1982), and at the phonological level, shown as the use of pitch rises...
The effect of pitch interplay on English-speaking learners of Mandarin

where native speakers would use falls, and vice versa (e.g., Backman, 1979; Lepetit, 1989). As part of ongoing efforts toward understanding L2 intonation acquisition, the current study extends cross-linguistic comparisons and investigates the effect of Mandarin intonation-tone pitch interplay on English speakers’ perception and production.

Shen (1990) categorizes different intonations in Mandarin into six groups: statements, lexically and grammatically unmarked yes-no questions (unmarked questions in this study), -ma particle questions, A-not-A questions, alternative questions, and wh-questions. In Mandarin, yes-no questions are typically produced by morphological means: the interrogative particle -ma is attached sentence-finally. In colloquial Mandarin, unmarked questions are produced with the surface word order identical to the corresponding statement sentence. This kind of question is called an unmarked question, as it does not have any syntactic question cue (Shen, 1990; Yuan, 2004). Shen (1990) measured fundamental frequency (F0) values at four points in sentences of different modalities: the starting point, highest peak, lowest trough, and ending point. Unmarked questions, like -ma questions, were found to have higher F0 values at all four points compared to statements. Shen’s findings corroborate earlier studies conducted by Ho (1976), and Tseng (1981), in that questions raise the overall tonal values of an utterance and statements lower them.

Unmarked yes-no questions also exist in English. For instance, “the Philadelphia train leaves at seven?” (Hirschberg & Ward, 1995, p. 407). The pitch typically rises to a high-final boundary tone in such unmarked questions, which conveys the meaning of non-finality and inquiry (Pierrehumbert & Hirschberg, 1990; Scherer, Ladd, & Silverman, 1984; Ohala, 1984; Wells, 2006). English speakers are also perceptually accustomed to hearing a rise in question modality and have a preference for it. In a study of intonation and emotion, Scherer et al. (1984) asked English-speaking listeners to rate how agreeable sentences were. The participants rated yes-no questions with a final fall as strongly challenging. In contrast, rising yes-no questions were rated agreeable and polite. The interaction between certain intonational elements in the contour (e.g., the final pitch value) and other linguistic properties of utterances (e.g., sentence type) seems quite categorical when perception is concerned. The question is, will English speakers’ preference of final rise in questions impede their learning of different intonation + final tone combinations in another language?

According to the language survey in Ultan (1978), most languages employ intonation alone to signal questions, although their specific formation strategies may differ. The survey finds that 38 of a sample of 53 languages use a terminal pitch rise while 18 use a higher overall F0. With regards to the two languages
under examination in the current study, English exemplifies the first case while Mandarin exhibits the latter.

As will be reviewed in more detail in the following section, due to the dual usage of pitch in Mandarin (a tone language), a complicated and potentially confusing overlap exists between intonation and phonemic tones (Bolinger, 1989; Vance, 1976). Many studies focus on acoustic realization and perceptual cues of the pitch interplay in native speech (e.g., Lee, 2005; Lin, 2004; Liu & Xu, 2005; Shen, 1990; Tseng, 1981; Yuan, 2011), and other studies on L2 speakers acquiring intonation and tone separately (Chen, 1974; Liang & van Heuven, 2007; Shen, 1989; Visceglia & Fodor, 2006; Yang & Chan, 2010); fewer studies report details about the simultaneous realization of intonation and tone in L2 speech. Those studies that do (e.g., Yang & Chan, 2010) only examine perceptual aspects and leave aside production. To further examine the much-neglected topic of pitch interplay in L2 acquisition, the current study focuses on both the perception and production of unmarked questions and statements. The two modalities only differ in intonation, and thus provide us with an ideal setting, since syntactical cues do not play a role.

1.2 Mandarin pitch interplay in L1 speech

Existing analyses have shown that Mandarin questions feature increased F0 across their global domain, raising the pitch of the overall utterance (DeFrancis, 1963; Ho, 1976; Tseng 1981; Shen, 1990). In particular, many studies have observed a local pitch rise starting at or near the end of the question sentence (Chao, 1968; Tseng, 2003; Zeng, Martin, & Boulakia, 2004; Lee, 2005; Liu & Xu, 2005). While it is still not clear when exactly the tone starts to rise in questions, the general consensus is that it does so in the sentence-final position. The overall data suggest that Mandarin questions increase the value of sentence-final tone when the rising intonation movement coincides with tonal direction (i.e., with a rising tone), while the tonal value tends to be neutralized when the two directions are opposing (i.e., with a sentence-final falling tone) (Ho, 1976; Shen, 1990; Zeng et al., 2004). A detailed analysis by Shen (1986) revealed that high-level Tone 1 (T1) loses its level contour and varies along with intonation: it moves upward in questions and downward in statements. Both T2 (a rising tone) and T3 (a low-dipping tone) rise with question intonation. T4 does not move upward but the pitch range is much larger in questions than in statements, reflected in a deeper terminal drop. In regards to whether the sentence-final tone somewhat influences the global intonational patterns, Kratochvil (1998) argued that the dual usage of pitch in Mandarin tends to blur the visibility of intonation to the point that the very existence of intonation has been questioned. However, Yuan (2004) found that the effect of tone on intonation is not significant. Lacking more evidence, I will leave the influence of tone
on intonation to future discussion. Taken together, unlike the canonical terminal rise cuing questions in English, two mechanisms have been argued to account for differences between statements and unmarked questions in Mandarin: an overall higher pitch contour for question intonation, combined with higher pitch values of sentence-final tones (Yuan, 2006).

The pitch interplay between intonation and tone raises a question of how native listeners interpret simultaneous pitch cues; that is, whether tone becomes indistinguishable from intonation. One view is that intonation has little effect on native Mandarin listeners’ tone perception (Connell, Hogan, & Rozsypal, 1983). A recent study by Yuan (2011), however, reported that Mandarin participants’ perception was indeed compromised by the interaction of sentential intonation and lexical tone (see similar findings on Cantonese speakers in Ma, Ciocca, & Whitehill, 2006). Yuan’s research found that native listeners’ statement (S) identification was more accurate than their unmarked question (Q) identification (98% vs. 80%), and that many Qs were misidentified as Ss. Further analysis showed that the sentence-final tone did not affect S identification but did affect Q identification. Question identification was easier if the sentence-final tone was T4 (a falling tone), whereas it was harder if it was T2 (a rising tone). T2 is not misinterpreted as “a mechanism of question intonation” for natives but rather causes “a greater bias toward statement intonation” (p. 4068). The author assumes that the combination of Q+T2 is more difficult because T2 “masks” the question intonation, since the pitch movements of T2 and the question coincide; however, it is “a mystery” why questions are easiest to identify with a final T4 (p. 4068). These results are rather unexpected given the universal tendency for a final rise to cue a question (Gussenhoven & Chen, 2000; Hadding-Koch & Studdert-Kennedy, 1964; House, 2003).

1.3 Mandarin pitch interplay in L2 speech

Yuan’s (2011) study inspires a series of questions. The strategies employed by native Mandarin listeners for resolving simultaneous pitch cues of intonation and tone seem quite surprising and counterintuitive, in particular that they identify a question more easily with a falling sentence-final tone — an environment in which intonation and tone conflict. One question arises: will English speakers learning Mandarin perceive an apparent terminal-rising pitch in Mandarin as the sentence-final tone or will it be interpreted as a mechanism of question intonation, as it would be in English?

Liang and van Heuven’s (2007) study suggests that the sentence-final rise is conducive to question identification. The results showed that all listeners, regardless of their L1 background (i.e., tone vs. non-tone languages), identified pitch-manipulated unmarked sentences with a terminal rise as questions, whereas sentences
with a terminal fall were identified as statements. However, this result should be treated carefully. Although there was no lexical/syntactic question marker in the stimuli (i.e., -ma), the computer-manipulated tokens were highly likely to present unnatural acoustic cues that interfered with listeners’ judgment. Also, the testing stimuli consisted of T1 only, in order to reduce the lexical tonal influence, so one cannot draw any conclusions about the interplay between intonation and tone. Examining combinations of all lexical tones and intonations, the study by Yang and Chan (2010) found that Mandarin unmarked questions were especially difficult for English speakers to identify when ending in T3 and T4 (both low-pitch offset tones, conflicting with Qs). Listeners also found statements difficult when they ended in T2 (a high-pitch offset tone, conflicting with Ss). Yang and Chan concluded that English-speaking learners were greatly affected by the conflicting pitch environment, in contrast to Mandarin speakers.

The study of Yang and Chan (2010) provides us with comprehensive results on L2 intonation and tone perception; what they did not examine is a much-needed set of comparable production data to create a more complete picture of how English speakers are affected by pitch interplay. One such set is provided in Visceglia and Fodor (2006), which examined how native English speakers produced Mandarin yes-no questions (with the -ma particle) and echo questions (without -ma) in comparison to native Mandarin speakers. The Mandarin group was found to raise the pitch within a question well before the sentence-final particle -ma. In contrast, native English speakers produced a final rise only on the -ma particle in Mandarin questions or the last syllable in echo questions, which strongly suggests that they were transferring the English interrogative final rise in their production of Mandarin. If the -ma particle cues question modality for English speakers, it is worth investigating whether they will still produce a similar final rise in question forms without any syntactical cues. Also, Visceglia and Fodor’s study did not include statements, making it difficult to draw a baseline for a comparison between statement and question intonation.

Following previous studies on pitch interplay, the current study sets out to investigate the results reported in Yang and Chan (2010) on L2 perception of intonation and tone. More important, this study also aims at providing a preliminary examination of how perceptual patterns are reflected in production patterns.

1.4 Research questions and predictions

Table 1 summarizes the specific association of different pitch contours in English and in Mandarin as discussed earlier; falling pitch indicates statements in English while it is associated with both statements and questions in Mandarin, when ending in different tones. Similarly, rising pitch indicates questions in English but can
also denote statements in Mandarin. Also included in Table 1 are the respective predictions. The general hypothesis, based on previous studies (e.g., Yang & Chan, 2010), is that English speakers will have more difficulty in perceiving pitch in environments where intonation and tone conflict. Furthermore, predictions state that such a difficulty will manifest in production: English speakers will be less successful in producing intonation and tone when the pitch directions are opposite than when they move in the same direction.

Table 1. Pitch contours in English and in Mandarin

<table>
<thead>
<tr>
<th>English Prosody</th>
<th>Mandarin Prosody</th>
<th>Compatibility</th>
<th>Prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statement (S)</td>
<td>Falling pitch</td>
<td>S + T4</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>S + T2</td>
<td>No</td>
</tr>
<tr>
<td>Question (Q)</td>
<td>Rising pitch</td>
<td>Q + T4</td>
<td>No</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Q + T2</td>
<td>Yes</td>
</tr>
</tbody>
</table>

2. Method

2.1 Material

The stimuli used in this study consisted of 64 sentences: 32 minimal pairs of statements and unmarked questions. None of the stimuli contained syntactic markers signalling their sentence modality; that is, all sentences could be interpreted as either a statement or a question based on intonation alone. The 64 utterances were divided into eight groups, each group containing eight utterances with the same number of syllables and the same grammatical structure. The eight sentences in each group represented all possible combinations of four lexical tones and two intonations. Each sentence was controlled for the tones of its last 2 or 3 syllables: T1+T1/2/3/4, T2+T1/2/3/4, T3+T1/2/3/4, T4+T1/2/3/4, T1+T1+T1/2/3/4, T2+T2+/T1/2/3/4, T3+T3+T1/2/3/4, T4+T4+T1/2/3/4.1 This design assured the inclusion of every possible combination of intonation and sentence-final tones. One female native Mandarin speaker recorded all the material (see the full list in Appendix). For example:

1. One reviewer pointed out that T2-T3 and T3-T3 are the same tone sequences because of Tone 3 sandhi. Because the focus of this paper is not on Tone 3 sandhi produced by English learners, and not even on tone production itself, our discussion of the results will not include Tone 3 sandhi production.
(1) Statement:
Míngtiān gāi Xiăo Zhāng kāi fēijī.
‘It is Xiao Zhang’s turn to fly the airplane tomorrow.’

(2) Question:
Míngtiān gāi Xiăo Zhāng kāi fēijī?
‘It is Xiao Zhang’s turn to fly the airplane tomorrow?’

2.2 Participants

Because of the rather extensive research results available on L1 speech (Liu & Xu, 2005; Yuan, 2011), a control group of Mandarin speakers was considered less necessary; instead, the focus was on L2 speakers’ performance. A group of ten Mandarin learners (F = 4, M = 6) with an average age of 24.8 years old participated in the current study. Prior to the experiment, a questionnaire was administered to elicit background information (e.g., age, gender, total years of Mandarin instruction, amount of in-country exposure, knowledge of additional languages). To assure within-group homogeneity, only participants with intermediate and above proficiency were recruited for this study.

The proficiency judgement was based on two criteria: learning time and accent rating. First, all participants had been learning Mandarin for at least one and a half years, with an average learning time of four years; nine of the ten considered themselves as above intermediate in terms of their overall proficiency in Mandarin. Second, a pre-task was conducted to better assess participants’ proficiency levels. Each participant was given four sentences in Mandarin that were also translated into English to ensure that they understood the meaning. The sentences were adopted from Wang and Shi (2010) and pre-recorded by the female native Mandarin speaker who recorded all perceptual materials. Participants heard the four sentences and were then recorded repeating the four sentences. Ratings of accent were assigned by two native Mandarin speakers on a scale from zero to ten, zero corresponding to native-like and ten to an extremely heavy accent. No participants received an accent rating higher than seven, and the average accent rating score across the two judges was 3.3/10. This result indicates that participants were perceived to be slightly accented in this study, where five represented the middle point. All participants were native North American English speakers.
(nine Canadians, one American). They were born into English-speaking families and began their Mandarin learning after puberty. None of them had experience learning another tonal language.

### 2.3 Procedure

The primary experiment used E-Prime 2.0 behavioral research software (Schneider, Eschman, & Zuccolotto, 2012) and consisted of three tasks. Each task used the same 64 sentences in random order and each sentence appeared once in each task. Task 1 asked participants to listen to one sentence at a time and choose the sentence intonation (statement vs. question). Task 2 asked participants to identify the sentence-final tone (T1, T2, T3 or T4) when listening to the sentences again. Task 3 was a production task that involved listening to a sentence and producing it with a different intonation; if the sentence heard was a statement, participants were asked to produce it as a question, and vice versa. This task was designed to examine what strategies participants were using to distinguish statements from questions in a relatively spontaneous speech. To reduce memory-related effects, each sentence was given on the screen in Task 3, and transcribed in pinyin with each tone mark added to its associated vowel, with the exception of the final syllable.

Before the primary experiment, a training session was given to familiarize the participants with the procedure. The training session was designed exactly the same way as the primary experiment but with fewer and different testing materials (eight sentences in each training task). The whole procedure took approximately 30 to 45 minutes. All participants received either an hourly Mandarin tutoring session or an honorarium of ten Canadian dollars upon completing the experiment.

### 3. Results

#### 3.1 Intonation perception

In Task 1, native English-speaking listeners were asked to identify whether the 64 Mandarin sentences were statements or questions. E-prime 2.0 collected the perceptual data. A total of 640 responses (10 subjects * 64 sentences) were generated and subjected to analysis. The results showed that the average intonation identification accuracy was 82.5%. Participants were more accurate at identifying

---

2. Two participants reported that they had a Cantonese-speaking mother and a Shanghainese-speaking father, respectively; however, they also reported that none of their parents used Mandarin at home.
S intonation than Q intonation (91.3% vs. 73.8%). Figure 1 below summarizes the results of intonation perception.

![Intonation perception accuracy](image)

**Figure 1.** Perception results for intonation  
Note: The graph indicates the participants’ accuracy in perceiving the two intonations (statements vs. questions) ending in different sentence-final tones. The higher the bar, the better the accuracy. Error bars: 95% Confidence Interval.

As expected, not all intonation + tone combinations were equally difficult to distinguish. The highest accuracy for sentences (Ss) was with T4 in sentence-final position (100%), and the highest accuracy for questions (Qs) was with T2 (90%). The lowest accuracy for Ss was with T2 (83.8%) and for Qs was with T1 (66.3%). The success in perceiving Ss when sentences end in T4 was in line with predictions, though the 100% accuracy was rather unexpected given that the group was entirely composed of L2 speakers. The effects of intonation and tone on the identification of intonation were analyzed using a Two-way Repeated-Measures ANOVA. The two independent variables were sentence type (S vs. Q) and tone (T1, T2, T3, T4). The dependent variable was average accuracy. Results showed that the main effect of sentence type was statistically significant \[F(1, 9) = 20.103, p < 0.002\], with S easier to identify than Q. The main effect of tone was not significant \[F(3, 27) = 1.803, p > 0.174\].

Because the interaction between sentence type and sentence-final tone was significant \[F(3, 27) = 6.695, p < 0.004\], multiple comparisons with Bonferroni corrections were conducted to further explore intonation and tone interactions. The results showed that the effect of all tones except T2 was significant for identifying intonation. Participants had better accuracy in identifying Ss than Qs when sentences ended with T1, T3, or T4 (all \(p < 0.029\)). Their intonation accuracy differences were not significant when sentences ended with T2 \[F(1, 9) = .849, p > 0.381\], with Qs easier to identify than Ss. That is, all tones except T2 facilitate the identification of S intonation. In statements, participants had significantly better
accuracy when the sentences ended with T4 than with T1 \((p < 0.023)\). In questions, participants had significantly better accuracy when the sentences ended with T2 than with T1 \((p < 0.011)\). The reason that the main effect of tone was not significant is likely that the effect of T2 was opposite in S and Q conditions: participants performed significantly worse at identifying intonation in S+T2 and significantly better at identifying intonation in Q+T2. This was expected since the pitch direction in S+T2 is conflicting but is compatible in Q+T2.

As one reviewer points out, it seems that L2 learners achieve such high accuracy with Q+T2 because they simply take a rising pitch contour as the signal of a question and then apply that theory everywhere. That is, learners actually do not have the real ability to distinguish different intonation types. This observation is borne out in learners’ great difficulty in identifying S+T2: they frequently misjudged sentences of this kind as questions because of the final rising. Moreover, the difficulty of perceiving Qs is also supported by the low accuracy in perceiving Q+T1 (rising intonation + high tone offset). The reason for this difficulty is likely that T1 is a level tone creating no final rise or final fall in the sentence, thereby making intonation ambiguous to judge. This finding further suggests that the English-speaking participants could not categorize stimuli as questions without a clear terminal-rising pitch, even if the sentence-final tone pitch was high in F0.

3.2 Tone perception

In Task 2, native English-speaking listeners were asked to identify the sentence-final tone in each sentence they were presented with. In all, the participants generated 640 responses (10 participants * 64 sentences) for analysis.

The results of Task 2 are illustrated in Figure 2. The overall tone identification accuracy was 80.6%, slightly lower than intonation accuracy (82.5%). Figure 2 shows that tone identification was more accurate for statements (Ss) than for questions (Qs) (88.6% vs. 72.7%). It also shows that participants experienced extreme difficulty identifying T3 and T4 in Qs, as seen in their dramatically lower accuracy rates: 45.2% and 68.9%, respectively. In contrast, the identification rates for other tones were above 80% in both the S and Q intonational contexts. The highest accuracy was observed for T4 in statements (95.1%), followed by T2 in questions (90.1%). This result was in keeping with findings from Task 1; that is, participants were most accurate in identifying the intonation of statements ending with T4 and questions ending with T2.

A Two-way Repeated-Measures ANOVA revealed that the two main effects (sentence type and tone) on the identification of tone were statistically significant. The statistical significance of the main effect of sentence type \([F(1, 9) = 17.465, p < 0.002]\) suggests that one sentence type made tone identification easier than
the other (i.e., S was easier than Q). The statistical significance of the main effect of tone \([F(3, 27) = 4.830, p < 0.024]\) indicates that not all tones were equally easy to identify. In addition, there was a significant interaction between sentence type and tone \([F(3, 27) = 14.005, p < 0.001]\). To further explore their interaction, multiple comparisons with Bonferroni corrections were conducted. The results showed that participants made significantly better judgements in identifying T3 \((p < 0.001)\) and T4 \((p < 0.007)\) in Ss than in Qs. The differences in tone judgment were found under Qs as well: T2 was found to elicit significantly better accuracy than T3 in Qs \((p < 0.014)\). Also, the identification accuracy between T1 and T3 was significantly different in Qs \((p < 0.035)\).

These results indicate that participants performed significantly worse in perceiving T3 and T4 in Qs than in perceiving other tones. This is to be expected because both T3 and T4 end with a low offset, which is in the opposite pitch direction to that of English. Together with the 100% accuracy of S+T4 and 90% accuracy of Q+T2 in Task 1, the perception evidence indicates that participants were more accurate at decoding intonation and tone information when their pitch moved in compatible directions, but less accurate in cases of conflicting pitch movements.

### 3.3 Intonation production

The perceptual results generally correspond to previous research, in that question intonation proved harder to identify than statement intonation (as in Yuan, 2011), and conflicting pitch environments impeded L2 listeners’ identification (as in Yang & Chan, 2010). The other research goal concerns English-speaking learners’ production of pitch interplay. To this end, the two native speakers who rated the participants’ accentedness were asked to rate the participants’ intonation.
As stated earlier, the tokens in Task 3 were the same as in Task 1 and Task 2: 32 statements and 32 questions differing only in intonation. The participants were instructed to complete the reverse-reproduce task: produce a statement if they heard a question, and a question if they heard a statement. The two judges were then presented with pairs of sentences (a statement and a corresponding unmarked question) and were explicitly asked to identify which was the statement and which was the question based on pitch movement. The first two participants’ data were excluded due to slight revisions to the stimuli-prompting slides following their data collection sessions, leaving eight participants’ data subject to analysis ($F = 4, M = 4$). Each participant produced a set of 64 sentences, yielding a total of 512 tokens (64 tokens * eight participants). The overall number of sentences leading to disagreements between the two judges was 105/512, placing inter-rater reliability at 79.5%: a fairly high level of consistency. A correlation analysis was conducted to compare the two judges’ ratings. Results showed that their ratings were strongly correlated [$r(8) = .951, p < 0.001$].

Intonation production accuracy was calculated by counting the intended intonations that were incorrectly rated by at least one judge. As shown in Table 2, the average accuracy across the eight participants was 56.3%, which was quite low considering their low accent ratings and their perceptual accuracy above 80%. The results showed that only 40.6% of intended Qs were identified as such while intended Ss were identified as such 71.5% of the time. A two-tailed T-test showed that the difference between Q identification and S identification was statistically significant ($p < 0.01$), suggesting that participants performed significantly worse in producing Qs than in producing Ss.

<table>
<thead>
<tr>
<th>Participant</th>
<th>Intended Qs identified as Qs</th>
<th>Intended Ss identified as Ss</th>
<th>Overall identification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average M</td>
<td>30.4%</td>
<td>82.8%</td>
<td>56.5%</td>
</tr>
<tr>
<td>Average F</td>
<td>51%</td>
<td>60%</td>
<td>56%</td>
</tr>
<tr>
<td>Average</td>
<td>40.6%</td>
<td>71.5%</td>
<td>56.3%</td>
</tr>
</tbody>
</table>

The failure to produce a distinction between Ss and Qs was especially interesting in the case of two male participants (P5 and P10) who were rated among the top 3 most fluent speakers by both judges, and yet produced fewer than 20% of the intended questions successfully. Their perceived speech fluency in terms of accent-edness did not give them any advantage in producing the modality distinction.

3. The tone was added to its carrying vowel instead of presenting it as a number after each syllable, according to P2’s suggestion. P1 and P2’s data were therefore excluded.
further examine their results I measured the F0 values at four points produced by P5 from a randomly chosen sentence pair (S vs. Q): starting point, valley, peak, and ending point. The F0 values of the four points were then compared to those produced by P3 (the most successful participant in the intonation production task, also male) for the same pair. The comparison here focuses only on global pitch range (i.e., no pitch values of final tones were involved). Figure 3a displays a clear difference between P3’s question (higher F0) and statement (lower F0). Figure 3b shows that, unlike P3, P5 did not produce much pitch variation in the two different sentence types: the F0 values at four points almost coincided in S and Q. Also, P5’s pitch range was significantly smaller than P3’s (50–180 Hz vs. 50–220 Hz).

As discussed above, a high global pitch contour is associated with Mandarin questions, whereas here a pattern of (a) a narrow pitch range and (b) no variation
The effect of pitch interplay on English-speaking learners of Mandarin

between intonations was found across English-speaking participants’ productions in this study. The average pitch value for each of the 64 sentences produced by all participants was obtained using a PRAAT script (Boersma & Weenink, 2013). The overall results are presented in Table 3 below.

Table 3. Mean F0 for each participant (Hz)

<table>
<thead>
<tr>
<th>Participant</th>
<th>Statement</th>
<th>Question</th>
<th>Mean Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>P3 (M)</td>
<td>145.16</td>
<td>170.63</td>
<td>25.47</td>
</tr>
<tr>
<td>P4 (M)</td>
<td>133.85</td>
<td>127.52</td>
<td>−6.33</td>
</tr>
<tr>
<td>P5 (M)</td>
<td>117.53</td>
<td>117.61</td>
<td>.08</td>
</tr>
<tr>
<td>P6 (F)</td>
<td>224.19</td>
<td>223.18</td>
<td>−1.01</td>
</tr>
<tr>
<td>P7 (F)</td>
<td>223.55</td>
<td>219.38</td>
<td>−4.55</td>
</tr>
<tr>
<td>P8 (F)</td>
<td>194.66</td>
<td>194.91</td>
<td>.25</td>
</tr>
<tr>
<td>P9 (F)</td>
<td>219.48</td>
<td>223.85</td>
<td>4.37</td>
</tr>
<tr>
<td>P10 (M)</td>
<td>137.61</td>
<td>130.13</td>
<td>−7.48</td>
</tr>
</tbody>
</table>

The results showed that only P3 had more than 25 Hz spans in producing different intonations. A two-tailed t-test was conducted to examine whether there was a significant difference between participants’ pitch range and that of the native speaker in the recordings. The results of female and male participants were collapsed together because only P3 had a clear pitch range distinction in producing the two intonations. Aside from P3, all other participants’ mean F0 differences between S and Q were significantly different from those of the native speaker \( t(6) = -5.775, p < 0.001 \). This difference explains their failure to make intonational distinctions, and affirms that native judgment is not based on impression, but on acoustic properties.

3.4 Tone production

In addition to intonation ratings, the two judges also identified the sentence-final tone to assess tone production accuracy. Of the 512 tokens (64 sentences * 8 participants), twenty that could not be judged due to partial sound loss were excluded. In all, the two judges identified 492 sentence-final tones. A Pearson correlation showed that both judges’ ratings of tones were significantly correlated \( r(247) = .61, p < 0.001 \). Tonal accuracy was calculated by counting the number of tones that failed to be correctly rated by at least one judge. There were 168 errors in total, making their average accuracy 65.9% (168/492), which is higher than intonation production accuracy (56.3%). Given the higher accuracy rates in compatible pitch environments shown in the perceptual tasks, English-speaking
participants were expected to produce T2 most often in Q, and T4 most often in S. This speculation was only partially supported: the two judges did identify T2 most frequently in Q, but they identified T4 least often in S.

Table 4 summarizes the tone errors in questions on which both judges agreed in their identifications. Both judges identified incorrect tones (i.e., the tones produced by mistake) as T2 most frequently (54 times), accounting for 50.9% of all misproductions in agreed questions. Also, when participants intended to make T4, native judges considered that they were producing T1 (20 times) and T2 (19 times).

In Ss, English-speaking participants were expected to produce T4 the most often. Unlike for Qs, however, the incorrect tones for Ss were more evenly distributed. As Table 5 shows, most of the mistaken tones in Ss were judged as T1 (i.e., an intended T2 was judged as T1 17 times). Also, when participants intended to

<table>
<thead>
<tr>
<th>Table 4. Tone errors in questions*</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Target</strong></td>
</tr>
<tr>
<td><strong>Judged As</strong></td>
</tr>
<tr>
<td>T1</td>
</tr>
<tr>
<td>T2</td>
</tr>
<tr>
<td>T3</td>
</tr>
<tr>
<td>T4</td>
</tr>
<tr>
<td>Total</td>
</tr>
</tbody>
</table>

* Note: Tones T1, T2, T3, and T4 along the x-axis represent the tonal targets that the participants were supposed to produce in agreed questions. On the y-axis, tones T1, T2, T3, and T4 represent the actual tones judged by the two judges. The “Total” column summarizes the number of tones that failed to be judged as the target tone. Where participants produced incorrect tones, they tended to use T2 to replace other tones.

<table>
<thead>
<tr>
<th>Table 5. Tone errors in statements*</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Target</strong></td>
</tr>
<tr>
<td><strong>Judged As</strong></td>
</tr>
<tr>
<td>T1</td>
</tr>
<tr>
<td>T2</td>
</tr>
<tr>
<td>T3</td>
</tr>
<tr>
<td>T4</td>
</tr>
<tr>
<td>Total</td>
</tr>
</tbody>
</table>

* Note: Tones T1, T2, T3, and T4 along the x-axis represent the tonal targets that the participants were supposed to produce in agreed statements. On the y-axis, tones T1, T2, T3, and T4 represent the actual tones judged by the two judges. The “Total” column summarizes the number of tones that failed to be judged as the target tone. Where participants produced incorrect tones, their tendency was evenly distributed.
make T1, the target tone (T1) was identified as other tones 18 times; when they were to make T4, the target tone (T4) was judged as another 10 times. The fact that participants misproduced T1 the most often leaves the hypothesis unsupported.

4. Summary and discussion

This study examined in detail how English speakers discriminate intonation and tone in Mandarin. Since the sentences differed only by virtue of their prosody, the task challenged participants’ ability to correctly perceive and produce Mandarin intonation and tone in compatible as well as conflicting pitch contexts. As predicted, a conflicting pitch environment portends more difficulty than a compatible one for English speakers in both perception and production. The findings can be summarized into five observations. First, participants performed better at perceiving intonation than tone (82.5% vs. 80.6%), while they performed better at producing tone than intonation (65.9% vs. 56.3%). Second, participants were more accurate in perceiving Ss than Qs (91.3% vs. 73.8%), and were also better at producing Ss than Qs (71.5% vs. 40.6%). Third, participants were most accurate at perceiving both intonation and tone when they were in compatible pitch movements (e.g., rising tone with rising question intonation), and less successful otherwise. Fourth, tone – in particular T2 – affected intonation identification. Intonation also affected tone identification, especially in Qs, which negatively affected the perception of T3 and T4. Finally, participants had a preference for producing T2 at the ends of questions (rising intonation + rising final tone), but they did not show the same kind of preference for producing T4 at the ends of statements (falling intonation + falling final tone).

4.1 Perceptual findings

The perceptual findings in this study are in line with previous research, showing that L2 learners resolve the conflict between intonation and tone cues differently from native Mandarin speakers. First, conflicting/compatible pitch environments affect L2 perception more seriously than they affect L1 perception. In this study, English speakers had particular difficulty processing sentences in which intonation and sentence-final tone moved in opposite directions; they were much more successful at identifying both intonation and tone in compatible pitch directions (also see Yang & Chan, 2010). Q+T2 (both rising) and S+T4 (both falling) were the easiest combinations for intonation identification among our English-speaking participants; Q+T3/4 (rising intonation + falling tone) and S+T2 (falling intonation + rising tone) were the most difficult. These findings contrast with those of
Yuan (2006, 2011), which showed that for native Mandarin speakers, intonation was the easiest to identify in Q+T4. Secondly, tone affects L2 intonation identification more significantly than it does L1 perception. For example, Yuan’s (2011) study found that for Mandarin listeners, sentence-final tone only affected Q identification; this study revealed that sentence-final tone affected perception of both sentence types. This is because participants performed significantly worse at identifying Ss with a sentence-final T2 but significantly better at identifying Qs with T2 (compared to other tones).

In this study, Q was found to be harder to identify than S, consistent with previous studies on L2 as well as on L1 perception (Yang & Chan, 2010; Yuan, 2011). An interesting finding is that English participants had difficulty identifying intonation in cases of Q+T1 (a high-level tone): such combinations were mostly considered statements. Previous studies (Pierrehumbert & Hirschberg, 1990; Scherer et al., 1984; Wells, 2006) have suggested that English speakers associate final rise with questions; the current study further reveals that not only can English speakers not distinguish sentence types when a final falling pitch is involved, but they also cannot determine a question without a clear terminal-rising pitch. That is, they systematically rely on terminal rising as a mechanism for Q association. A flat contour, even if it is high in F0, still cues S rather than Q for English listeners.

4.2 Production findings

4.2.1 Intonation production
The perceptual difficulty of Qs is reflected in participants’ production. The production results in this study indicate that English speakers had more difficulty producing intelligible questions than statements (40.6% vs. 71.5%). This production difference may be accounted for by differing perceptual accuracies (91.3% vs. 73.8% in Task 1). That is, when one participant produced a question, there might be two scenarios: (1) the participant heard a statement then correctly judged it as such and produced a question with the correct intonation; and (2) the participant heard a statement but judged it a question, and meant to produce a statement, but ended up producing a question. The second possibility cannot be completely ruled out in this study; however, if it is the case, it further confirms that perpetual difficulties can transfer to production problems.

More likely, however, the differing production accuracy is due to participants’ flat and narrow pitch range when producing sentences (e.g., Chen, 1974; Visceglia & Fodor, 2006). As Barlow (1998) points out, learners tend to underutilize the prosodic means available to them in the L2, resulting in their intonation patterns being flatter and narrower than the target norms.
This study not only found that English speakers exhibit small pitch ranges, but also established the relationship between acoustic measurements and native judgements. The acoustic analyses discovered that, except for P3, all participants had significantly different pitch ranges from that of the native speaker, and those speakers failed to make their intonation – especially in unmarked questions – intelligible. This observation of L2 speakers exhibiting smaller pitch range is consistent with other studies on the acquisition of English intonation patterns, by, for instance, German speakers of English (Jilka, 2007) and Spanish speakers of English (Barlow, 1998). The current data provide preliminary evidence for extending cross-linguistic comparison, and point towards small pitch range as a universal pattern in L2 speech.

4.2.2 Tone production

The perception and production data are complementary in terms of the effect of compatible versus conflicting pitch movements. In the perception tasks, participants had more difficulty perceiving both intonation and tone when their pitch movements did not coincide. This difficulty was reflected in production: participants showed a definite preference for T2 at the end of a question (compatible pitch movements). This suggests that the L1 English speakers in this study are transferring the English interrogative final rise to the final syllable of their Mandarin interrogatives, regardless of the final tone (Visceglia & Fodor, 2006).

The current study further suggests that this transfer occurs in unmarked questions and substantially affects intelligibility, given that most of the final rises were mistakenly identified as a lexical tone (T2) by the native Mandarin speakers. Interestingly, participants did not show a similar preference for T4 (a falling tone) at the ends of statements; rather, native judges identified more T1s than other tones here. That English speakers make more errors with T1 is also found in Shen (1989), and in Zhang and Yuan (2011). It is likely that T4 behaves similarly in English contours, in which it marks a word boundary or the end of a statement (Shen, 1989). Broselow et al. (1987) showed that when T4 occurred in the position where English-speaking subjects were accustomed to hearing it in L1 (e.g., word-finally), those subjects were quite successful in their perception of it. This word-level success is extended to the sentence level in the current study: the English speakers did not produce more T4s at the ends of statements than they did at the ends of questions. In any case, the fact that T4 is not the preferred tone in S (as T2 is in Q) suggests that the interaction between intonation and tone is stronger in Q than in S. This concurs with Yuan’s (2011) findings that intonation-tone interactions are stronger in Qs than in Ss for native Mandarin speakers.

This study has also shown that English learners of Mandarin were better at perceiving intonation than tone but better at producing tone than intonation. There are
similar results in Visceglia and Fodor (2006), finding that the tone production of Mandarin learners is more accurate than their intonation production. They argue that this may be attributed to attentional focus, in that English-speaking learners of Mandarin may have focused on local phonetic features (i.e., tone), and therefore may not have paid attention to the global pitch height of Ss and Qs. Another possibility is that learners rely on lexical information when listening to stimuli (Lieberman, 1965), and as such, have better tone judgment given lexical information embedded in the context. However, it was difficult to determine whether the participants had already learned the specific words examined in the current study.

4.3 Pedagogical implication

The intonation production issue is highly likely because of minimal teaching of intonation. As many scholars note (e.g., Celce-Murcia et al., 1996; Grice & Bauman, 2007; Gut, Trouvain, & Barry, 2007), intonation is seldom taught systematically and still plays a minor role in the L2 classroom, even though it is particularly difficult for non-native learners to master. Derwing and Munro (2005) address this issue, stating that it is due to lack of knowledge: teachers are often left to rely on their own intuitions and given little direction on how to teach intonation. For example, my observation is that many Mandarin instructors emphasize the dip-rising contour of T3 in teaching, while simply asking students to produce a low onset for T3 would make their pronunciation much more accurate.

In this study, several participants stated that they were not used to using intonation to signal a question in Mandarin because their instructors explicitly instructed them never to do so, nor to use the terminal-rise pattern. Apparently, their teachers were aware of the influence stemming from the L1 (the final rise), and learners were instructed to only form -ma or A-not-A questions. If the aim of language teachers is to enable language learners to acquire the prosody of the target language to an adequate extent, the instruction goal should range from minimal communicative abilities to a near-native language competence (Gut et al., 2007). Thus, although unmarked questions are less frequent in real language use, it is by no means less important to teach this question form for communicative as well as language-development purposes. In this study, even for those intermediate-advanced learners, producing unmarked questions still presented a large amount of communicative misunderstanding to native speakers, affecting the learners’ communicative competence and creating a potential obstacle to achieving native-like fluency. The results from this study explicitly show that instructors can simply teach students to raise the global pitch contour when intending a question, and their pronunciation will be much more intelligible.
5. Conclusion

In the current study, English-speaking learners of Mandarin manifested two kinds of native-language influence, especially when producing unmarked questions: a phonetic influence (i.e., the relatively smaller pitch range), and a phonological influence (i.e., the systematic replacing of other tones with T2 when intending a question). The consequence of the narrower pitch range is that the intonation patterns in the L2 production were not recognizable by native speakers as the intended intonation. In addition to L1 influence, the difficulty of producing appropriate intonation contours could be attributed to the lack of intonation instruction in the Chinese language classroom. This study provides instructional direction for Chinese teachers by raising the awareness of intonation as a significant component of Chinese language pedagogy.

References


doi: 10.1515/9783110853469


doi: 10.1007/11939993_7


### Appendix. Training session

- 小张喜欢猫. (xiao3 zhang1 xi3 huan1 mao1.)
- 小张喜欢猫? (xiao3 zhang1 xi3 huan1 mao1?)

- 小王没钱买菜. (xiao3 wang2 mei2 qian2 mai3 cai4.)
- 小王没钱买菜? (xiao3 wang2 mei2 qian2 mai3 cai4?)

- 小张有五把雨伞. (xiao3 zhang1 you3 wu3ba3 yu3san3.)
- 小张有五把雨伞? (xiao3 zhang1 you3 wu3ba3 yu3san3?)

- 小王最爱吃面. (xiao3 wang2 zui4 ai1 chi1 mian4.)
- 小王最爱吃面? (xiao3 wang2 zui4 ai1 chi1 mian4?)

### Testing stimuli

#### Group 1

- 明天该小张开车. (Ming2 tian2 gai1 xiao3 zhang1 kai1 che1.)
- 明天该小张开车? (Ming2 tian2 gai1 xiao3 zhang1 kai1 che1?)
明天该小张开球 (Ming2 tian2 gai1 xiao3 zhang1 kai1 qiu2.)
明天该小张开球? (Ming2 tian2 gai1 xiao3 zhang1 kai1 qiu2?)

明天该小张开讲. (Ming2 tian2 gai1 xiao3 zhang1 kai1 jiang3.)
明天该小张开讲? (Ming2 tian2 gai1 xiao3 zhang1 kai1 jiang3?)

明天该小张开会. (Ming2 tian2 gai1 xiao3 zhang1 kai1 hui4.)
明天该小张开会? (Ming2 tian2 gai1 xiao3 zhang1 kai1 hui4?)

**Group 2**

小王不喜欢名车. (xiao3 wang2 bu4 xi3 huan1 ming2 che1.)
小王不喜欢名车? (xiao3 wang2 bu4 xi3 huan1 ming2 che1?)

小王不喜欢名牌. (xiao3 wang2 bu4 xi3 huan1 ming2 pai2.)
小王不喜欢名牌? (xiao3 wang2 bu4 xi3 huan1 ming2 pai2?)

小王不喜欢名表. (xiao3 wang2 bu4 xi3 huan1 ming2 biao3.)
小王不喜欢名表? (xiao3 wang2 bu4 xi3 huan1 ming2 biao3?)

小王不喜欢名片. (xiao3 wang2 bu4 xi3 huan1 ming2 pian4.)
小王不喜欢名片? (xiao3 wang2 bu4 xi3 huan1 ming2 pian4?)

**Group 3**

明天小张要打工. (ming2 tian1 xiao3 zhang1 yao4 da3 gong1.)
明天小张要打工? (ming2 tian1 xiao3 zhang1 yao4 da3 gong1?)

明天小张要打球. (ming2 tian1 xiao3 zhang1 yao4 da3 qiu2.)
明天小张要打球? (ming2 tian1 xiao3 zhang1 yao4 da3 qiu2?)

明天小张要打狗. (ming2 tian1 xiao3 zhang1 yao4 da3 gou3.)
明天小张要打狗? (ming2 tian1 xiao3 zhang1 yao4 da3 gou3?)

明天小张要打架 (ming2 tian1 xiao3 zhang1 yao4 da3 jia4.)
明天小张要打架? (ming2 tian1 xiao3 zhang1 yao4 da3 jia4?)
**Group 4**

小王明天要上班.  
(xiao3 wang2 ming2 tian1 yao4 shang4 ban1.)

小王明天要上班?  
(xiao3 wang2 ming2 tian1 yao4 shang4 ban1?)

小王明天要上学.  
(xiao3 wang2 ming2 tian1 yao4 shang4 xue2.)

小王明天要上学?  
(xiao3 wang2 ming2 tian1 yao4 shang4 xue2?)

小王明天要上网.  
(xiao3 wang2 ming2 tian1 yao4 shang4 wang3.)

小王明天要上网?  
(xiao3 wang2 ming2 tian1 yao4 shang4 wang3?)

小王明天要上课.  
(xiao3 wang2 ming2 tian1 yao4 shang4 ke4.)

小王明天要上课?  
(xiao3 wang2 ming2 tian1 yao4 shang4 ke4?)

**Group 5**

小张没见过飞机师.  
(xiao3 zhang1 mei2 jian4 guo4 fei1 ji1 shi1.)

小张没见过飞机师?  
(xiao3 zhang1 mei2 jian4 guo4 fei1 ji1 shi1?)

小张没见过飞机员.  
(xiao3 zhang1 mei2 jian4 guo4 fei1 ji1 yuan2.)

小张没见过飞机员?  
(xiao3 zhang1 mei2 jian4 guo4 fei1 ji1 yuan2?)

小张没见过飞机场.  
(xiao3 zhang1 mei2 jian4 guo4 fei1 ji1 chang3.)

小张没见过飞机场?  
(xiao3 zhang1 mei2 jian4 guo4 fei1 ji1 chang3?)

小张没见过飞机票.  
(xiao3 zhang1 mei2 jian4 guo4 fei1 ji1 piao4.)

小张没见过飞机票?  
(xiao3 zhang1 mei2 jian4 guo4 fei1 ji1 piao4?)

**Group 6**

小张要用留言箱.  
(xiao3 zhang1 yao4 yong4 liu2 yan2 xiang1.)

小张要用留言箱?  
(xiao3 zhang1 yao4 yong4 liu2 yan2 xiang1?)

小张要用留言牌.  
(xiao3 zhang1 yao4 yong4 liu2 yan2 pai2.)

小张要用留言牌?  
(xiao3 zhang1 yao4 yong4 liu2 yan2 pai2?)
小张要用留言笔。  (xiao3 zhang1 yao4 yong4 liu2 yan2 bi3.)
小张要用留言笔?  (xiao3 zhang1 yao4 yong4 liu2 yan2 bi3?)

小张要用留言信。  (xiao3 zhang1 yao4 yong4 liu2 yan2 xin4.)
小张要用留言信?  (xiao3 zhang1 yao4 yong4 liu2 yan2 xin4?)

**Group 7**

小王负责演讲厅。  (xiao3 wang2 fu4 ze2 yan3 jiang3 ting1.)
小王负责演讲厅?  (xiao3 wang2 fu4 ze2 yan3 jiang3 ting1?)

小王负责演讲台。  (xiao3 wang2 fu4 ze2 yan3 jiang3 tai2.)
小王负责演讲台?  (xiao3 wang2 fu4 ze2 yan3 jiang3 tai2?)

小王负责演讲稿。  (xiao3 wang2 fu4 ze2 yan3 jiang3 gao3.)
小王负责演讲稿?  (xiao3 wang2 fu4 ze2 yan3 jiang3 gao3?)

小王负责演讲赛。  (xiao3 wang2 fu4 ze2 yan3 jiang3 sai4.)
小王负责演讲赛?  (xiao3 wang2 fu4 ze2 yan3 jiang3 sai4?)

**Group 8**

小王喜欢运动装。  (xiao3 wang2 xi3 huan1 yun4 dong4 zhuang1.)
小王喜欢运动装?  (xiao3 wang2 xi3 huan1 yun4 dong4 zhuang1?)

小王喜欢运动鞋。  (xiao3 wang2 xi3 huan1 yun4 dong4 xie2.)
小王喜欢运动鞋?  (xiao3 wang2 xi3 huan1 yun4 dong4 xie2?)

小王喜欢运动品。  (xiao3 wang2 xi3 huan1 yun4 dong4 pin3.)
小王喜欢运动品?  (xiao3 wang2 xi3 huan1 yun4 dong4 pin3?)

小王喜欢运动裤。  (xiao3 wang2 xi3 huan1 yun4 dong4 ku4.)
小王喜欢运动裤?  (xiao3 wang2 xi3 huan1 yun4 dong4 ku4?)
摘要

本实验探讨了汉语语调和句末声调对英语为母语的汉语学习者的影响。实验结果表明当语调和声调的基频走势一致时，学习者在听辨语调和句调时正确率都较高。在无句法标记的疑问句的输出中，二语学习者普遍音域较窄，句末音高突出。本实验再次强调了母语的影响以及语调在课堂教学中的重要性。

关键词: 无句法标记的疑问句, 声调, 语调, 声调和语调的输出

Author’s address

Shan Luo
Linguistics Department
University of Victoria
3800 Finnerty Road
Victoria, BC, V8P 5C2
Canada
shanluo@uvic.ca