



PERCEPTION OF MANDARIN INTONATION

Jiahong Yuan

Cornell University

ABSTRACT

This study investigates how tone and intonation, and how focus and intonation, interact in intonation type (statement vs. question) identification. A perception experiment was conducted on a speech corpus of 1040 utterances. Sixteen listeners participated in the experiment. The results reveal three asymmetries: statement and question intonation identification; effects of the sentence-final Tone2 and Tone4 on question intonation identification; and effects of the final focus on statement and question intonation identification. These asymmetries suggest that: 1. Statement intonation is a default or unmarked intonation type whereas question intonation is a marked intonation type. 2. Question intonation has a higher prosodic strength at the sentence final position. 3. There is a tone-dependent mechanism of question intonation at the sentence-final position.

1. INTRODUCTION

Chinese is a tonal language. There are four lexical tones in Mandarin Chinese, referred to as Tone1, Tone2, Tone3 and Tone4. The F_0 contours of the tones in isolation are high level, rising, low dipping (or just low), and falling, respectively. Both tone and intonation use F_0 as a primary cue. How, then, do tone and intonation interact with each other? For example, if a sentence ends with a falling tone, how can question intonation, which normally has a rising end in English and other non-tonal languages, be realized? Is it difficult for question intonation to be realized on such a sentence? On the other hand, for question intonation that has a rising tone at the end, how does a listener know that the utterance is a question? Or how does a listener tease apart the tone and intonation information from the surface rising end? By investigating the degrees of perception confusion when intonation types are realized on different tonal sequences in both focused and unfocused environments, we may find some clues as to how tone and intonation interact and how intonation is realized. This study follows this strategy.

A perception experiment was conducted on a speech corpus of 1040 utterances. Section 2 and Section 3 describe the corpus and the methods of the perception experiment respectively. The results are presented in Section 4 in two categories: intonation type identification and effects of focus on intonation type identification. Implications of the results are discussed in Section 5.

2. CORPUS

A corpus of 130 sentences was designed. The sentences, all of which contain eight syllables, are minimal sets contrasting *Intonation type* (Statement or Question), *Presence of a focus* (yes or no), *Focus position* (beginning, middle or end of a sentence), *Tone of the focused syllable* (Tone1, Tone2, Tone3 or Tone4) and *Tone of the last syllable* (Tone1, Tone2, Tone3 or Tone4). For example¹:

1. Li3bai4wu3 Luo2Yan4 yao4 mai3 maol.
Friday Luo2Yan4 will buy cat
“Luo2Yan4 will buy a cat Friday.”
[statement, last syllable is Tone1, no focus]
2. Li3bai4wu3 Luo2yan4 yao4 mai3 maol?
Friday Luo2yan3 will buy cat
“Luo2Yan4 will buy a cat Friday?”
[question, last syllable is Tone1, no focus]
3. Li3bai4wu3 Luo2Yan4 yao4 mai3 **yang2**.
Friday Luo2Yan4 will buy **goat**
“(not a cat, not a deer) Luo2Yan4 will buy
a **goat** Friday.”
[statement, last syllable is Tone2,
focus at the end, focused syllable is Tone2]

For those sentences that were designed to have a focus, a context was provided to introduce the focus by excluding alternatives, following the semantics of focus proposed in [1]. For example, the context for “Luo2Yan4

¹ The examples are transcribed in *Pinyin* romanization with a number at the end of each syllable indicating the tone of the syllable.

will buy a **goat** Friday” is “not a cat, not a deer”. Each context was bracketed and placed before the sentence.

Eight native Mandarin speakers, four male and four female, took part in the recording. The speakers were asked to speak the sentences as a question if there is a question mark at the end and a statement if there is a period at the end. In addition, if there is a context the speakers were asked to read the context silently and then speak out the following sentence in a way suitable to the context.

A total of 1040 utterances (130 sentences for each speaker for eight speakers) were obtained.

3. METHODS

3.1. Experimental procedure

A perception experiment for identifying intonation type was conducted on the 1040 utterances recorded in the database. The experiment consisted of one pre-test training block and two test blocks. The pre-test training block contained 23 utterances. Each of the two test blocks contained 520 utterances.

The utterances were played to the listeners in a randomized order through a speaker in a quiet room, using E-Prime software installed on a PC. The Inter Stimulus Intervals (ISIs) were 2500ms. There was a five minute break between the two test blocks.

The listeners were asked to listen to the utterances carefully and record their response to each utterance on the answer sheets provided to them. They were asked to write a ‘?’ at the end of a sentence if the utterance was a question and write a ‘.’ at the end if it was a statement.

Sixteen listeners, 8 female and 8 male, participated in the perception experiment. All of them are native Mandarin speakers.

3.2. Intonation type identification ratio

‘Intonation type identification ratio’ is used to measure how well an intonation type was recognized by the listeners [2]. Each of the 1040 utterances has an ‘intonation type identification ratio’, which is defined as the ratio of the ‘correct’ responses (statement or question) over the total 16 responses by the listeners. For example, if utterance A was produced as a question by a speaker and 12 of the 16 listeners perceived utterance A as a question, then the intonation type identification ratio of utterance A is 0.75 (12/16).

A high ‘intonation type identification ratio’ of an utterance suggests that the intonation type of the utterance is easy to recognize. If an intonation type of an utterance is difficult to recognize, suggested by a low identification ratio, the intonation type is confusable to listeners.

4. RESULTS

4.1. Intonation type identification

As presence of focus may affect the identification of an intonation type, the ‘no focus’ utterances were used to analyze intonation type identification.

The mean intonation type identification ratios across all the speakers plus and minus two standard errors of each mean, which construct a 95 percent confidence interval around the mean, are drawn in Figure 1.

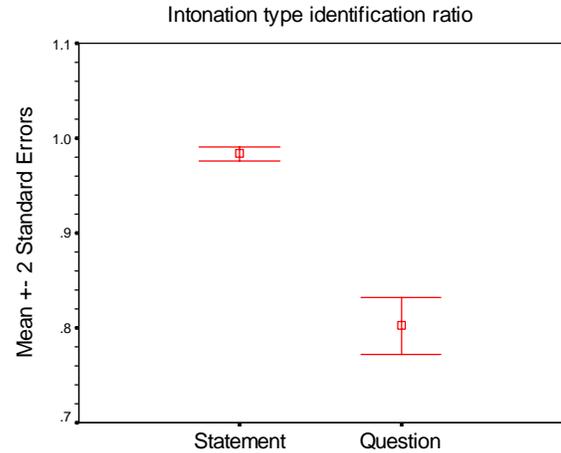


Figure 1: Mean intonation type identification ratios of statement and question intonation across all speakers

Clearly, statement intonation is better identified than question intonation.

Using a method called *simple main effects* [3], we can compare each pair of *tone of the last syllable* under each *intonation type*. The results are provided in Table 1.

Table 1: *Tone of the last syllable difference*

Intonation type	(I) Tone of the last syllable	(J) Tone of the last syllable	Mean Difference (I-J)	Sig.
Statement	Tone1	Tone2	-.003	.91
	Tone1	Tone3	-.021	.30
	Tone1	Tone4	.003	.88
	Tone2	Tone3	-.018	.41
	Tone2	Tone4	.006	.80
	Tone3	Tone4	.024	.25
Question	Tone1	Tone2	.074	.00
	Tone1	Tone3	-.022	.27
	Tone1	Tone4	-.120	.00
	Tone2	Tone3	-.096	.00
	Tone2	Tone4	-.193	.00
	Tone3	Tone4	-.098	.00

We can see from Table 1 that for statement intonation no pair of *tone of the last syllable* has significant difference, for question intonation, however, the difference is significant for all pairs except the pair of Tone1 and Tone3.

The mean differences between Tone2 and the other tones and those between Tone4 and the other tones are reformatted and listed below:

Tone2 – Tone1: -.074	Tone4 – Tone1: .120
Tone2 – Tone3: -.096	Tone4 – Tone2: .193
Tone2 – Tone4: -.193	Tone4 – Tone3: .098

Clearly, Tone2 is lower than the other tones and Tone4 is higher than the other tones. This suggests that question intonation identification is better/easier if the tone of the last syllable in the utterance is Tone4 whereas it is worse/harder if the tone of the last syllable is Tone2.

4.2. Effects of focus on intonation type identification

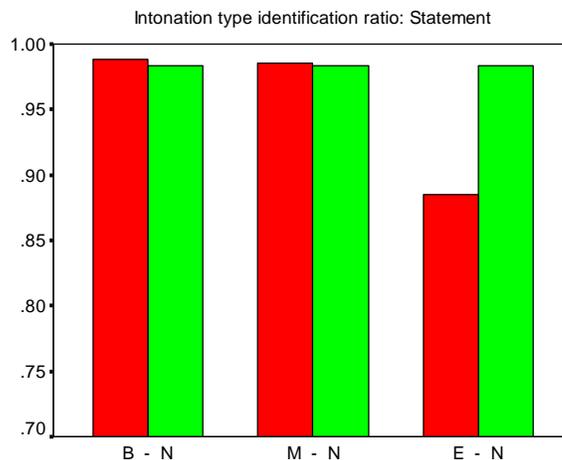
This section studies the effects of focus on intonation type identification. All utterances, including both ‘no focus’ and ‘having a focus’, were used. The difference between having a focus at the beginning, middle, or end of a sentence and having no focus for each intonation type was investigated using *simple main effects*. Table 2 displays the results.

Table 2: *Focus position difference*

Intonation type	(I) Focus position	(J) Focus position	Mean Difference (I-J)	Sig.
Statement	Beginning	No focus	.006	.68
	Middle	No focus	.004	.76
	End	No focus	-.101	.00
Question	Beginning	No focus	.013	.37
	Middle	No focus	.098	.00
	End	No focus	.097	.00

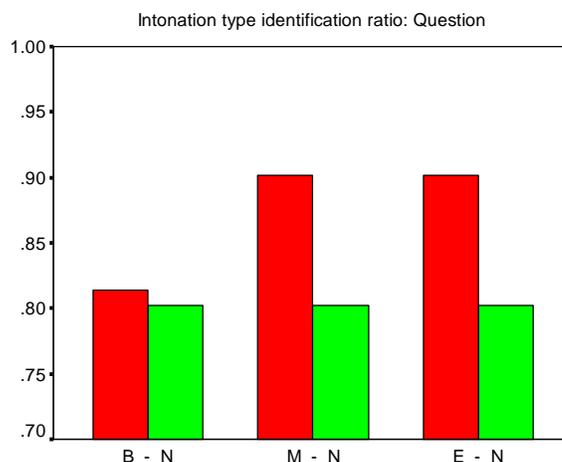
We can draw the following conclusions from Table 2:
 1. A focus at the beginning or middle position of a sentence does not affect the identification of statement intonation whereas a focus at the end of an utterance makes statement intonation more difficult to identify.
 2. A focus at the middle or the end of a sentence makes question intonation easier to identify whereas a focus at the beginning does not affect the identification of question intonation.

Figures 2 and 3 illustrate the effects of focus position on statement and question intonation identification respectively.



B: ‘having a focus’ at the beginning; M: ‘having a focus’ at the middle; E: ‘having a focus’ at the end; N: no focus.

Figure 2: *Mean intonation type identification ratios of ‘no focus’ vs. ‘having a focus’: statement intonation*



B: ‘having a focus’ at the beginning; M: ‘having a focus’ at the middle; E: ‘having a focus’ at the end; N: no focus.

Figure 3: *Mean intonation type identification ratios of ‘no focus’ vs. ‘having a focus’: question intonation*

5. DISCUSSION

The results presented in Section 4 reveal three interesting asymmetries: statement and question intonation identification; effects of the final Tone2 and Tone4 on question intonation identification; and effects of the final focus on statement and question intonation identification.

The asymmetry of statement and question intonation identification manifests in two ways: First, statement intonation is easier to identify than question intonation; second, the tone of the last syllable does not affect statement intonation identification but it does affect question intonation identification. The intonation

identification test was a forced choice test: the listeners must identify the intonation type of each utterance as either a statement or a question. That question intonation identification was less accurate means that many question intonation utterances were identified as statements. This suggests that statement intonation is a default or unmarked intonation type. That is, listeners fall back to this option when there is not enough information suggesting ‘question’, which is also supported by the fact that the tone of the last syllable does not affect statement intonation identification. Question intonation is, however, a marked intonation type. It can only be identified if listeners actually hear the ‘question’ features/mechanisms. The fact that the tone of the last syllable affects question intonation suggests that the ‘question’ features/mechanisms conflict with some tonal features but not others.

The second asymmetry revealed by the perception experiment is of the effects of the sentence-final Tone2 and Tone4 on question intonation identification: question intonation identification is better/easier if the tone of the last syllable in the utterance is Tone4 whereas it is worse/harder if the tone of the last syllable is Tone2. Tone4 is a falling tone and Tone2 is a rising tone. Therefore the asymmetry can also be stated as follows: In sentence-final position, question intonation is easier to identify on a falling tone and more difficult to identify on a rising tone. This finding clearly suggests an intonation model where the realization of intonation type is sensitive to the tonal identity, as proposed by Shih in [4]. This aspect is not captured by most of the Chinese intonation models in the literature. Although the issue of interaction of tone and intonation has been addressed in these models, none of them is capable of explaining this asymmetry. Most of these models assume that sentence intonation contour is a high level phenomenon. The intonation contour representing each intonation type would have been chosen by high level information such as semantics, pragmatics, and speaker intention, and the sentential intonation pattern would have been determined before the selection of the tones. Representative views of Chinese question intonation include treating it as rising grid [5], rising curve [6], separately functioned top line and base line [7], high boundary tone [8], and raised pitch level and exaggerated pitch range [9]. Most of these models have no provision to allow on-line tone and intonation interaction. The asymmetry of the effects of the sentence-final Tone2 and Tone4 on question intonation identification suggests, however, that there is a tone-dependent mechanism of question intonation at the sentence final position.

The third asymmetry is that a focus at the end of a sentence makes statement intonation harder to identify but makes question intonation easier to identify. This finding is consistent with the strength mechanism of question intonation found in [9]: Question intonation has higher

prosodic strengths in the sentence final syllables. Both question intonation and a final focus have a higher strength at the sentence final position. Therefore, presence of both in a sentence will make it easier for the listeners to identify the higher strength mechanism, which is an indicator of question intonation to the listeners in the intonation type identification test. If there is a focus at the end of a statement, the higher strength of the last focused tone may be misinterpreted as a mechanism of question intonation for some listeners. Therefore more statements were identified as questions if focus was presented in final position. It would be interesting to look at the results when listeners identify both the intonation type and the focus position at the same time. I expect that, in such a situation, a focus at the end of a sentence will probably not make statement intonation harder to identify and question intonation easier to identify, because the higher strength of the final focus is less likely to be misidentified as a mechanism of question intonation when both focus position and intonation type are identified.

6. ACKNOWLEDGEMENT

I thank Chilin Shih, Abigail Cohn, Greg Kochanski, Amanda Miller-Ockhuizen and Mats Rooth for their suggestions and comments.

7. REFERENCES

- [1] Rooth, M., *Association with Focus*, Ph.D. dissertation, University of Massachusetts, Amherst, 1985.
- [2] J. Yuan and C. Shih, “Confusability of Chinese intonation,” *Proceedings of Speech Prosody 2004*, Nara, Japan, pp. 131-134, 2004.
- [3] Winer, B.J., D.R. Brown, and K.M. Michels, *Statistical Principles in Experimental Design*, McGraw-Hill, New York, 1991.
- [4] C. Shih, “Tone and intonation in Mandarin,” *Working Papers of the Cornell Phonetics Laboratory* 3, pp. 83-109, 1988.
- [5] E. Gårding, “Speech act and tonal pattern in Standard Chinese: constancy and variation,” *Phonetica* 44, pp. 13-29, 1987.
- [6] Shen, X., *The Prosody of Mandarin Chinese*, University of California Press, Berkeley, 1989.
- [7] J. Shen, “Hanyu yudiao gouzao he yudiao leixing,” [Intonation structure and intonation types of Chinese], *Fangyan* 3, pp. 221-228, 1994.
- [8] S. Peng, M. Chan, C. Tseng, T. Huang, O. Lee, and M. Beckman, “A Pan-Mandarin ToBI,” <http://people.cohums.ohiostate.edu/chan9/MToBI.htm>
- [9] J. Yuan, C. Shih, and G.P. Kochanski, “Comparison of declarative and interrogative intonation in Chinese,” *Proceedings of Speech Prosody 2002*, Aix-en-Provence, France, pp. 711-714, 2002.