Confusability of Chinese Intonation

*Jiahong Yuan & Chilin Shih**

Department of Linguistics, Cornell University
*Department of East Asian Languages and Cultures and Department of Linguistics,
University of Illinois at Urbana-Champaign
jy55@cornell.edu cls@uiuc.edu

Abstract

Do lexical tones interfere with the realization of intonation types? Given that tone and intonation both use F_0 as a primary cue, can a listener reliably identify statements and questions when some of the channel capacity is taken up by lexical tones? We study this issue through a perception test on a carefully designed and obtained intonation corpus on Mandarin Chinese. Our study shows the following: 1. Statement intonation is easier to recognize than question intonation; 2. the sentence-final tone does not affect statement intonation recognition; 3. question intonation is easier to recognize if the sentence-final tone is falling whereas it is harder to recognize if the sentence-final tone is rising. Implications of the results for the modeling of Chinese intonation are discussed.

1. Introduction

Do lexical tones interfere with the realization of intonation types? Given that tone and intonation both use F_0 as a primary cue, can a listener reliably identify statements and questions when some of the channel capacity is taken up by lexical tones? Mandarin Chinese has four lexical tones: tone1 (high), tone2 (rising), tone3 (low) and tone4 (falling). Both tone and intonation are primarily realized on F_0. This may cause conflicts when producing intonation on a sequence of tones. We study this issue through a perception test on a carefully designed and obtained intonation corpus on Mandarin Chinese.

By investigating the degrees of perception confusion when intonation types are realized on different tonal sequences, we may find some clues as to how tone and intonation interact and how intonation is realized. This paper follows this strategy. A similar methodology has been adopted by Yuan et al. in their study of emotion in Chinese and proved successful [1].

Our previous study on tone modeling [2] shows that the distinction between Mandarin question and statement intonation patterns becomes increasingly pronounced at the end of the sentence. This result is consistent with the cross-linguistic study of question intonation [3] where it is found that language may employ several possible strategies to differentiate statement and question intonation, such as final rise, final high pitch, late peak, and late prominence, many of them involve high or strong gestures near the end of the sentence. The ToBI-based intonation school [4] also uses a phrasal tone and a final boundary tone to represent the difference between question and statement intonation. All these studies point to the importance of the sentence final locality to the realization of intonation type. Hence in this study we test the interaction of intonation and the tone of the final syllable.

Section 2 and section 3 describe the corpus and the perception test on the corpus respectively. Section 4 reports the results of the perception test, which show the varying difficulty of Chinese intonation identification. Implications of the results for the modeling of Chinese intonation are discussed in section 5.

2. Corpus design and recording

A corpus of 64 sentences was designed. The experimental design includes eight sentence types. The sentences, all of which contain eight syllables, are minimal sets contrasting *Intonation type* (*Statement* and *Question*) or *Tone of the last syllable* (tone1, tone2, tone3 and tone4). For example[1]:

1. Li3bai4wu3 Luo2Yan4 yao4 mai3 mao1.
   Friday Luo2Yan4 will buy cat
   "Luo2Yan4 will buy a cat Friday."
   [statement, last syllable is tone1]

2. Li3bai4wu3 Luo2yan4 yao4 mai3 mao1?
   Friday Luo2yan3 will buy cat
   "Luo2Yan4 will buy a cat Friday?"
   [question, last syllable is tone1]

3. Li3bai4wu3 Luo2Yan4 yao4 mai3 yang2.
   Friday Luo2Yan4 will buy goat
   "Luo2Yan4 will buy a goat Friday."
   [statement, last syllable is tone2]

4. Li3bai4wu3 Luo2Yan4 yao4 mai3 yang2?
   Friday Luo2Yan4 will buy goat
   "Luo2Yan4 will buy a goat Friday?"
   [question, last syllable is tone2]

Eight native Mandarin speakers, four male and four female, took part in the recording. All of them came to U.S. from Mainland China less than three years ago. A total of 512 utterances (64 sentences per speaker by eight speakers) were obtained.

The sentences were presented one by one on a computer display to the speakers in a randomized order. The timing of sentence presentation was controlled by the speakers. By pressing any key on a keyboard a speaker could replace the sentence on the display with another one. The speakers were asked to speak the sentences as a question if there is a question

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

mark at the end and as a statement if there is a period at the end.

The recording was done in a soundproof recording booth in the Cornell Phonetics Lab. The speech was recorded on a DAT recorder and was digitized at the sampling rate of 44.1 KHz, and was subsequently downsampled to 22050Hz on a computer.

3. Perception test

A perception test on the 512 utterances was conducted. Sixteen listeners, 8 female and 8 male, participated in the perception test. All of them came to U.S. from Mainland China less than three years ago and speak standard Mandarin Chinese.

The utterances were played to the listeners in a randomized order through a speaker in a quiet room, using E-Prime software [5]. The inter-stimulus intervals were 2500ms. The listeners were asked to listen to the utterances carefully and judge the intonation type of each utterance, choosing either 'question' or 'statement' on the answer sheets provided to them.

4. Results

4.1. Intonation type identification ratio

We use 'intonation type identification ratio' to measure how well an intonation type is recognized by listeners. Each of the 512 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 indicates that the intonation type in question was often correctly identified on that sentence, whereas a low intonation type identification ratio indicates a poor recognition record**.**

With SPSS software [6], a general linear model procedure was utilized to test the effects of *Speaker*, *Intonation type* and *Tone of the last syllable* on intonation type identification ratio. The results are listed in Table 1. It is clear that all three factors and all the interactions between them are statistically significant (Sig. < .05).

Table 1: *Results of the General Linear Model Test*

Factor	F	Sig.
Speaker	24.852	.000
Intonation type	273.891	.000
Tone of the last syllable	10.997	.000
*Speaker * Intonation type*	35.832	.000
*Speaker * Tone of the last syllable*	3.311	.000
*Intonation type * Tone of the last syllable*	12.206	.000
*Speaker * Intonation type * Tone of the last syllable*	3.145	.000

Dependent variable: Intonation type identification ratio

Variance of the speakers, which says that some speakers were more able to produce proper intonation than the others, is not a primary concern of this paper. The effects of *Intonation type* and *Tone of the last syllable* on intonation identification will be discussed below, with an emphasis on what are common to all the speakers.

4.2. Intonation type

Table 2 lists the mean intonation type identification ratios of *Statement* and *Question* for each speaker.

A Paired-Samples T Test on data in Table 2 shows that the difference between *Statement* and *Question* is statistically significant (Sig. < .05). We therefore can draw a conclusion that *Statement* is easier to identify than *Question*. Figure 1 illustrates the difference graphically. In Figure 1, the mean intonation 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.

Table 2: *Mean intonation type identification ratios of statement and question for each speaker.*

Speaker	*Statement*	*Question*
S1	.9902	.9336
S2	.9863	.8848
S3	.9785	.9629
S4	.9824	.7480
S5	.9961	.7285
S6	.9375	.9902
S7	1.0000	.6094
S8	.9961	.5586
Total	**.9834**	**.8020**

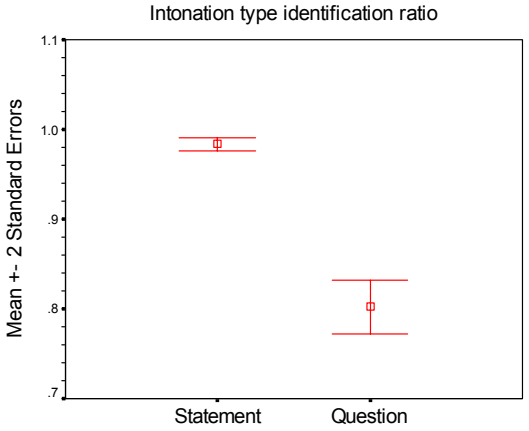

Figure 1: *Mean intonation type identification ratios of Statement and Question across all speakers.*

4.3. Tone of the last syllable

Table 3 and Table 4 list the mean intonation type identification ratios for each tone of the last syllable. Table 3 is of *Statement* and Table 4 is of *Question.*

From Table 3 we can see that *Tone of the last syllable* does not have a significant effect on *Statement* identification for all the speakers. The results of Paired-Samples T Tests on data in Table 3 back up this conclusion, showing that for *Statement* no pair is significantly different from each other.

Now we turn to question intonation. From Table 4 we can see that the identification ratio number of tone4 is higher than that of any other tone for all the speakers. This conclusion can also be verified by the results of Paired-Samples T Tests on data in Table 4, which show that the differences between tone1 and tone4, tone2 and tone4 and tone3 and tone4 are significant (Sig. < .05) whereas the others are not.

It is also worth noting that in Table 4 the identification ratios of tone2 for speakers 5 and 7 (.3750 and .2708 respectively) are much lower than those of the other tones. We can hence draw a conclusion that some speakers' question intonation is very difficult to recognize if it is realized on sentences ending with a tone2.

Figure 2 and Figure 3 illustrate the differences of the four tones for *Statement* (Table 3) and *Question* (Table 4) respectively. Again, the mean intonation type identification ratios across all the speakers plus and minus two standard errors of each mean are drawn in these figures.

Table 3: *Mean intonation type identification ratio for each tone of the last syllable: Statement.*

Speaker	*Tone of the last syllable*			
	Tone1	Tone2	Tone3	Tone4
S1	.9922	.9688	1.0000	.9911
S2	.9922	.9688	.9943	.9821
S3	.9219	.9896	1.0000	1.0000
S4	.9453	1.0000	.9943	.9911
S5	1.0000	1.0000	.9943	.9911
S6	.9766	.9063	.9943	.8304
S7	1.0000	1.0000	1.0000	1.0000
S8	.9844	1.0000	1.0000	1.0000
Total	**.9766**	**.9792**	**.9972**	**.9732**

Table 4: *Mean intonation type identification ratios for each tone of the last syllable: Question.*

Speaker	Tone of the last syllable			
	Tone1	Tone2	Tone3	Tone4
S1	.8906	.9271	.9489	.9643
S2	.8047	.9375	.8580	.9732
S3	.9297	.9583	.9659	1.0000
S4	.7031	.7813	.6591	.9107
S5	.7344	.3750	.7841	.9375
S6	.9844	.9688	1.0000	1.0000
S7	.5781	.2708	.7330	.7411
S8	.6328	.4479	.4830	.6875
Total	**.7822**	**.7083**	**.8040**	**.9018**

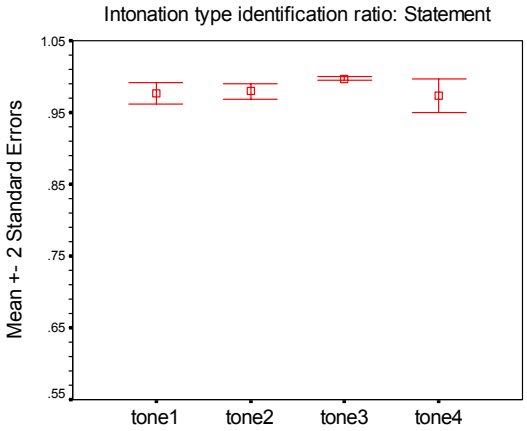

Figure 2: *Mean intonation type identification ratios on Tone of the last syllable for Statement.*

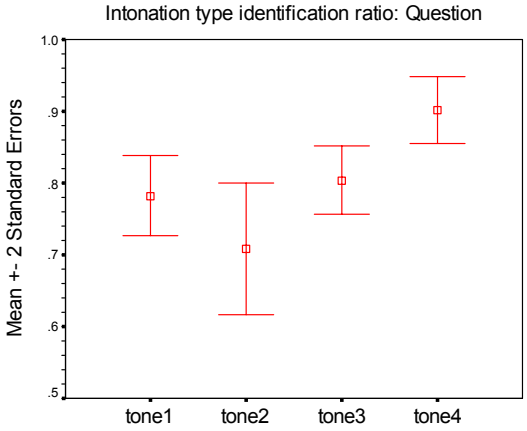

Figure 3: *Mean intonation type identification ratios on Tone of the last syllable for Question.*

4.4. Summary

To summarize the analysis above draws the following conclusions:

1. *Statement* is easier to identify than *Question*;
2. *Tone of the last syllable* does not affect the identification of the *Statement* intonation;
3. *Tone of the last syllable* affects the identification of the *Question* intonation: First, *Question* is easier to identify on sentences ending with a tone4 than those ending with the other tones; second, identification of some speakers' question is very difficult if the sentence ends with a tone2.

5. Discussion

Our study on the varying difficulty of Chinese intonation perception reveals two interesting asymmetries. The implications of them for the modeling of Chinese intonation will be discussed in the following.

5.1. Intonation type asymmetry

The first asymmetry is of *Statement* and *Question* identification. It manifests in two aspects: First, *Statement* is easier to identify than *Question*; second, *Tone of the last syllable* does not affect *Statement* identification but it does affect *Question* identification.

Question intonation is realized with exaggerated gestures in comparison with statement intonation. The pitch range is wider, and the final tone is higher and stronger than what is found in statements. If the exaggerated gestures are lacking, or are weakly executed, the signal becomes ambiguous and the listeners tend to choose the statement interpretation. One possible explanation is that the statement intonation is the default intonation type, which occurs with higher frequency than question intonation, so the listeners fall back to this option when there is no enough information suggesting the opposite.

5.2. Question intonation identification

The second asymmetry revealed by the study is of the effects of the sentence-final tone2 and tone4 on *Question* identification: On the one hand, *Question* is easier to identify on sentences ending with a tone4 than ending with the other tones; on the other hand, for some speakers *Question* is very difficult to identify on sentences ending with a tone2. Tone4 is a falling tone and tone2 is a rising tone in Chinese. Therefore the asymmetry can also be stated as follows: At the sentence-final position question intonation is easier to identify on a falling tone whereas harder to identify on a rising tone.

In the native speakers' intuition, question intonation in Chinese is rising but not falling, no matter its surface form has a rising end or a falling end (which mainly depends on the tone of the last syllable). Because the general directions of the gestures of question intonation and tone2 are the same, the effort of making a question intonation may be masked by the rising gesture of tone2. This might explain why at the sentence-final position question intonation is harder to identify on a rising tone.

Why question intonation is easier to identify on a falling tone, however, is a mystery. The combination of question intonation with tone1 and tone3 creates some modification in the tone shape. Final tone 1 (high) has an overall rising shape, final tone3 (low) obtains a distinctive rising tail. The combination of question intonation with tone2 (rising) and tone4 (falling), however, does not change the tone shape. Tone2 is still rising and tone4 is still falling. It is hence reasonable to expect that at the sentence-final position question intonation is easier to identify on tone1 and tone3 than on tone4, which is however, contrary to our finding.

Our finding clearly suggests an intonation model where the realization of intonation type is sensitive to the tonal identity, as proposed by Shih in [7]. 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 solving the mystery. 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 raised pitch level [8], rising grid [9, 10], rising curve [11], separately functioned top line and base line [12, 13], high boundary tone [14], and exaggerated and wide pitch range [2]. Most of these models have no provision to allow on-line tone and intonation interaction. Our findings in this paper suggest that the realization and perception of intonation type interact with lexical tone. We can isolate gestures that are relevant to the realization of question intonation, such as final high boundary tone, exaggerated effort and expanded pitch range toward the end of the sentence, however, how they are realized depends on the lexical tones.

6. Acknowledgement

We thank Abigail Cohn, Greg Kochanski, Amanda Miller-Ockhuizen and Mats Rooth for their wonderful suggestions and comments.

7. References

[1] Yuan, J.; Shen, L.; Chen, F., 2002. The Acoustic Realization of Anger, Fear, Joy and Sadness in Chinese. *Proceedings of ICSLP 2002*. Denver, Colorado, 2025-2028.

[2] Yuan, J.; Shih, C; Kochanski, G. P., 2002. Comparison of Declarative and Interrogative Intonation in Chinese. *Proceedings of Speech Prosody 2002*. Aix-en-Provence, France, 711-714.

[3] Gussenhoven, C.; Chen, Aoju, 2000. Universal and language-specific effects in the perception of question intonation. In *Proceedings of ICLSP 2000*, Beijing.

[4] Pierrehumbert, J. B., 1980. *The Phonology and Phonetics of English Intonation*. Ph.D. dissertation, MIT.

[5] Schneider, W.; Eschman, A.; Zuccolotto, A, 2002. *E-Prime User's Guide*. Psychology Software Tools, Inc.

[6] SPSS inc., 2001. *SPSS 11.0 for Windows*.

[7] Shih, C., 1988. Tone and Intonation in Mandarin. *Working Papers of the Cornell Phonetics Laboratory* 3, 83-109.

[8] DeFrancis, J. F., 1963. *Beginning Chinese*. New heaven: Yale University Press.

[9] Gårding, E., 1985. Constancy and variation in Standard Chinese tonal patterns. *Lund University Working Papers 28, linguistics-phonetics*, 19-51.

[10] Gårding, E., 1987. Speech act and tonal pattern in Standard Chinese: constancy and variation. *Phonetica* 44, 13-29.

[11] Shen, X., 1989. *The Prosody of Mandarin Chinese.* University of California Press.

[12] Shen, J., 1985. Beijinghua shengdiao de yinyu he yudiao[Pitch range of tone and intonation in Beijing dialect]. In *BeijingYuyin Shiyanlu,* Lin, T.; Wang L. (ed.). Beijing: Beijing University Press.

[13] Shen, J., 1994. Hanyu yudiao gouzao he yudiao leixing[Intonation structure and intonation types of Chinese]. *Fangyam* 3, 221-228.

[14] Peng, S. et al., 2000. A Pan-Mandarin ToBI. http://deall.ohio-state.edu/chan.9/MToBI.htm