Title: Perception of intonation in Mandarin Chinese

Author: Jihaong Yuan, University of Pennsylvania

Running title: Perception of intonation in Mandarin Chinese

Address:

    Jiahong Yuan

    609 Williams Hall, University of Pennsylvania

    Philadelphia, PA 19104

Email:

    jiahong@ling.upenn.edu

ABSTRACT

There is a tendency across languages to use a rising pitch contour to convey question intonation and a falling pitch contour to convey a statement. In a lexical tone language such as Mandarin Chinese, rising and falling pitch contours are also used to differentiate lexical meaning. How, then, does the multiplexing of the F_0 channel affect the perception of question and statement intonation in a lexical tone language? This study investigated the effects of lexical tones and focus on the perception of intonation in Mandarin Chinese. The results show that lexical tones and focus impact the perception of sentence intonation. Question intonation was easier for native speakers to identify on a sentence with a final falling tone and more difficult to identify on a sentence with a final rising tone, suggesting that tone identification intervenes in the mapping of F_0 contours to intonational categories and that tone and intonation interact at the phonological level. In contrast, there is no evidence that the interaction between focus and intonation goes beyond the psychoacoustic level. The results provide insights that will be useful for further research on tone and intonation interactions in both acoustic modeling studies and neurobiological studies.

Perception of intonation in Mandarin Chinese

I. INTRODUCTION

In a tone language such as Mandarin Chinese, F_0 serves as the primary acoustic cue for both lexical tones and sentence intonation. How do they interact with each other? Do lexical tones interfere with the realization and perception of intonation types? We study this issue through a perception experiment. By investigating how lexical tones influence the perception of intonation types, we may find some clues as to how tone and intonation interact and how intonation is perceived in a tone language.

The interaction between tone and intonation has been a central problem in acoustic studies of intonation in Mandarin Chinese. Chao (1968) described the interaction using two metaphors: *"The best answer is to compare syllabic tone and sentence intonation with small ripples riding on large waves (though occasionally the ripples may be "larger" than the waves). The actual result is an algebraic sum of the two kinds of waves."* (Chao 1968: 39); and: *"Stretching the [elastic transparent] sheet or letting it shrink will vary the pitch range of tones; doing the same horizontally will make the syllables longer or shorter"* (Chao 1968: 26). These ideas were widely adopted in early modeling studies of Mandarin intonation. Representative models of Chinese question intonation include treating the intonation as a rising grid (Gårding 1987), rising curve (Shen 1989), separately functioning top line and base line (Shen 1992), or high boundary tone (Peng et al. 2005). In these models, intonation is insensitive to lexical tones and is determined before tone selection. In recent years, quantitative modeling of Chinese intonation, in which the interaction between tone and intonation was quantified through

simulating surface F_0s, has been investigated (Yuan et al. 2002, Fujisaki et al. 2005, Ni and Hirose 2006, Prom-on et al. 2008). In these studies, the interaction was still largely treated as tone-independent. For example, Yuan and colleagues (2002) studied the difference between statement and question intonation in Mandarin Chinese by building detailed mathematical models of F_0 using Stem-ML (Soft Template Mark-up Language, Kochanski and Shih 2003). That study reported that the "diverse" difference between question and statement intonation in Mandarin Chinese can be accounted for by two consistent mechanisms for question intonation: an overall higher phrase curve and higher strength values of sentence final tones. Both of these mechanisms are tone-independent.

The interaction between tone and intonation is also an interesting issue in intonation perception. Listeners across languages, both tonal and non-tonal, tend to classify contours with a high peak, late peak, or terminal rise as questions, while they classify contours with a low peak or terminal fall as statements (Hadding-Koch and Studdert-Kennedy 1964, Gosy and Terken 1994, Gussenhoven and Chen 2000, House 2003). A widely accepted hypothesis is that such quasi-universal characteristics in the relationship between intonational form and meaning are innate and derived from biologically determined codes (Gussenhoven 2004). The frequency code (Ohala 1983, Gussenhoven 2002), for example, implies that a raised F_0 is a marker of submissiveness or non-assertiveness and hence of a question. Seemingly conflicting or similar acoustic cues of tone and intonation may cause perception confusion. For question intonation with a falling tone at the end, it is not clear how the conflict between the cues of tone and intonation is resolved. For question intonation with a rising tone at the end, how does a listener tease apart the tone and intonation information from the apparent rise at the end

of a question?

Despite many years of research, the principles underlying intonation processing are not well understood (reviewed in Vaissière 2005). Studdert-Kennedy and Hadding-Koch (1973) investigated auditory and linguistic interactions in the perception of intonation contours. These authors proposed that the perceptual process involves different stages (auditory, phonetic, phonological, etc.) and that there is feedback from higher to lower levels that may serve to correct or verify earlier decisions. Grabe and colleagues (2003) studied the native-language effect on the perception of intonation contours. Their results suggested that the perception of intonation begins with the activation of universal auditory mechanisms that separate falling from rising contours and that the output of the universal auditory process is then molded by the listener's experience with his or her native language. Other perceptual experiments have also demonstrated that the mapping between the acoustic signal and the intonational categories is phonologized (Chen et al. 2004, Makarova 2001). Liang and van Heuven (2007) compared L1 and L2 learners' perception of Chinese tone and intonation. They concluded that F_0 is primarily perceived at the lexical level for listeners of tone languages but at the sentence level for listeners of non-tone languages. They also suggested that lexical tone is the primary factor in reducing the sensitivity to pitch cues at the sentence level for listeners of a tone language. The neural basis of speech prosody, including tone and intonation, has been studied extensively (Baum and Penn 1999, Doherty et al. 2004, Gandour 2007). Although there is no clear consensus on what neural substrates underlie the perception of tone and intonation, it is generally accepted that for native speakers, tones are processed predominately in the left hemisphere, which is sensitive to the linguistic levels of

processing, while intonation is processed bilaterally with a preference for the right hemisphere, which is sensitive to lower-level acoustic processing. Gandour and colleagues (2004) concluded that there is a close interaction between the two hemispheres, writing that "*Chinese tone and intonation are best thought of as a mosaic of multiple local asymmetries that allows for the possibility that different regions may be differentially weighted in laterality depending on language-, modality-, and task-related features*" (Gandour et al. 2004: 350-351). Unlike acoustic studies, most perceptual and neurobiological studies of tone and intonation have focused on the effect of the linguistic knowledge possessed by native speakers or the differences between tone and intonation, but not the interaction between the two. It is therefore still unclear whether the perception of intonation is sensitive to tone, and at what level(s) tone and intonation interact with each other in perception.

This study aims to achieve a better understanding of the interaction between tone and intonation in Mandarin Chinese by investigating how lexical tones affect intonation perception, expanding our earlier study on this issue (Yuan and Shih 2004). The effect of focus on intonation perception will also be examined for two reasons. First, previous studies have found that the difference between statement and question intonation in Mandarin Chinese is more pronounced towards the end of the utterance (Tsao 1967, Ho 1977, Yuan et al. 2002), and this difference can be modeled as higher strengths of sentence final tones for question intonation (Yuan et al. 2002). If question intonation is conveyed by strengthening final syllables, then what is the difference between "plain" question intonation and question intonation in which the final syllable(s) are focused, given that a focused syllable has an expanded pitch range and therefore is also

strengthened (Gårding 1983, Eady and Cooper 1986, Shih 1988, Jin 1996, Xu 1999, Xu and Xu 2005, Chen and Gussenhoven 2008)? Furthermore, what is the difference between "plain" question intonation and statement intonation with a focus at its end? Second, the interactions between focus and intonation and between tone and intonation may have different properties. Liu and Xu (2005) investigated whether listeners could detect both focus and question intonation in the same utterance. Their results showed that listeners could identify both in most cases, indicating that F_0 variations related to the two functions are encoded in parallel and transmitted simultaneously. By comparing the effects of tone and focus on intonation perception, we can determine whether tone and focus interact with intonation differently.

The experiment tests Chinese listeners' ability to identify question and statement intonation on sentences that differ in tones and focus. The null hypothesis is that listeners can identify statements and questions with equal probability for sentences that differ in lexical tones and focus positions.

II. METHOD

A. Stimuli

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. A systematic corpus was created for the study. The corpus consisted of 130 sentences, which were minimal pairs of statement and question intonation. The last tone of each sentence was one of the four tones. Sixty-six of

the 130 sentences also had a narrow focus at the initial, middle, or final position of the sentence. For these sentences, some context was provided to introduce the focus, following the alternative semantics of focus (Rooth 1985; In this theory, focus is associated with a set of alternatives in the context). For example, the context to introduce a focus on 'goat' in the sentence "Luo2Yan4 will buy a **goat** Friday" is "(not a cat, not a deer)." The context was placed in brackets before the target sentence. Below are examples of two minimal pairs in the corpus (the sentences are transcribed in *Pinyin* romanization with a digit at the end of each syllable indicating the tone of the syllable):

Li3bai4wu3 Luo2Yan4 yao4 mai3 mao1.
Li3bai4wu3 Luo2Yan4 yao4 mai3 mao1?
"Luo2Yan4 will buy a cat Friday. / ?"
[The last syllable is Tone1, and there is no focus.]

(bu2shi4 mao1, bu2shi4 lu4) Li3bai4wu3 Luo2Yan4 yao4 mai3 **yang2**.
(bu2shi4 mao1, bu2shi4 lu4) Li3bai4wu3 Luo2Yan4 yao4 mai3 **yang2**?
"(not a cat, not a deer) Luo2Yan4 will buy a **goat** Friday. / ?"
[The last syllable is Tone2, and there is a focus at the end.]

Eight native Mandarin speakers, four male and four female, took part in the recording. The speakers were graduate students at Cornell University. The sentences were presented one by one in Chinese characters on a Computer display to the speakers. The set of sentences with no focus were presented first, followed by the sentences with focus. Within each block the sentences were presented in a randomized order. The speakers were asked to speak the sentences as a question when there was a question mark at the end and as a statement when there was a period at the end. In addition, if there was a context introducing a narrow focus in the target sentence, the speakers were asked to read the context silently and then to read the target sentence out loud in a way that was

appropriate for the context. The recording produced 1,040 utterances in total. The speakers were paid for their participation.

B. Procedure

A perception experiment was conducted using the 1,040 recorded utterances in the database. The utterances were played to the listeners in randomized order through a speaker in a quiet room, using E-Prime software installed on a PC. The inter-stimulus interval (ISI) was 2.5s. The listeners were asked to identify the intonation type of each utterance as a statement or question and to record their responses on the answer sheets provided to them. On the answer sheets, the texts (in Chinese characters) of the utterances played to the listeners were listed with no punctuation at the end. The listeners were asked to write a "?" at the end of a sentence if the utterance was a question and to write a "." at the end if it was a statement. The 1,040 sentences were printed on 11 answer sheets. The listeners took a short break after each answer sheet was completed. The experiment took approximately 90 minutes.

C. Subjects

Sixteen listeners, eight female and eight male, participated in the perception experiment. The listeners were graduate students at Cornell University who had not received any formal training in linguistics or speech technology prior to the experiment. Eight of the listeners were the same speakers who had made the recordings (the perceptual experiment was conducted one week after their recordings). This made it possible to explore whether different speakers have different intonational

strategies/phonologies. If this is the case, then we should expect the following results: 1. A speaker identifies intonation of his/her own speech better than intonation of others' (S identifies S's intonation better than S identifies others' intonation). 2. A speaker identifies intonation of his/her speech better than others do (S identifies S's intonation better than others identify S's intonation). The results, however, did not support the hypothesis that different speakers have different intonational phonologies (see more discussion in Yuan 2004). All of the subjects had come to the U.S. from northern China (Mandarin dialect area) within the three previous years and spoke standard Mandarin Chinese. They were between 25 and 35 years old and reported no hearing problems. The listeners were paid for their participation.

III. ANALYSIS

A. Effects of final tone on intonation-type identification

"Identification rate," defined as the ratio of the "correct" responses over the total of 16 responses for each utterance, is used to measure how well an intonation type was recognized by the listeners. A response to an utterance is "correct" if and only if it is identified as the same intonation type as what the speakers were asked to produce. A high "identification rate" suggests that the intonation type of the utterance is easy to recognize, and a low identification rate suggests that it confuses listeners.

As the presence of focus may affect the identification of an intonation type, the "No focus" utterances were used to analyze the effects of final tone on intonation-type identification. The "No focus" utterances refer to the utterances produced when only a

target sentence (designed as "No focus," without a focus-inducing context) was given to the speakers. Overall, statement intonation had both a higher mean of identification rates and smaller variation. The mean identification rates for statement and question intonation were 0.98 and 0.80, respectively, and the standard deviations for the two intonation types are 0.05 and 0.24, respectively.

With SPSS software, a general linear model (ANOVA) procedure was used to test the effects of *speaker*, *intonation type* and *tone of the last syllable* on intonation type identification. The results showed that the factors of *intonation type*, *tone of the last syllable* and the interaction between them were statistically significant ($p < .05$). The factor *speaker* was not significant but the interaction between *speaker* and *intonation type* and the interaction of all three factors were significant.

Further analysis showed that for statement intonation no pair of options for *tone of the last syllable* produced a significant difference. For question intonation, however, the difference was significant for all pairs except the pair of Tone1 and Tone3. The mean differences of identification rate between Tone2 and the other tones and those between Tone4 and the other tones for question intonation are listed below:

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

(All differences are significant at the .05 level.)

Clearly, the identification rate for Tone2 was lower than the other tones, and the identification rate for Tone4 was higher than the other tones. This result suggests that

question identification is better/easier if the tone of the last syllable in the utterance is Tone4 (falling tone) whereas it is worse/harder if the tone of the last syllable is Tone2 (rising tone).

Figures 1 and 2 illustrate the effects of *tone of the last syllable* on intonation-type identification for statement and question intonation, respectively.
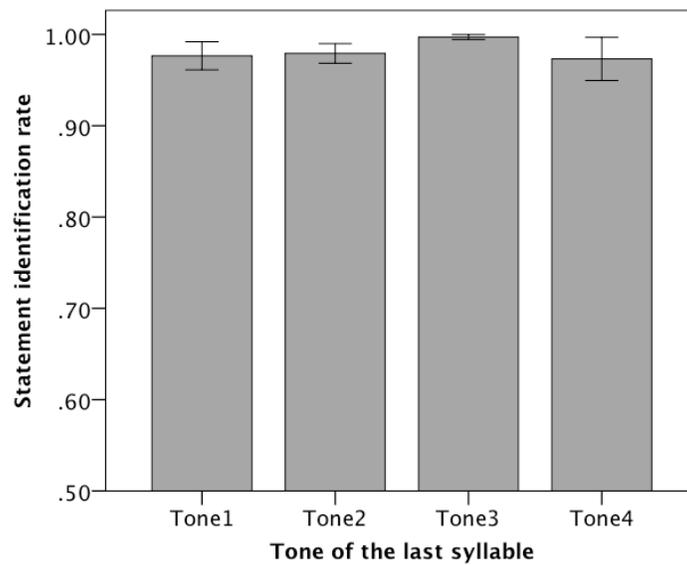

Fig. 1. Mean identification rate of statement intonation, grouped by tone of the last syllable. The differences among the tones are not significant.
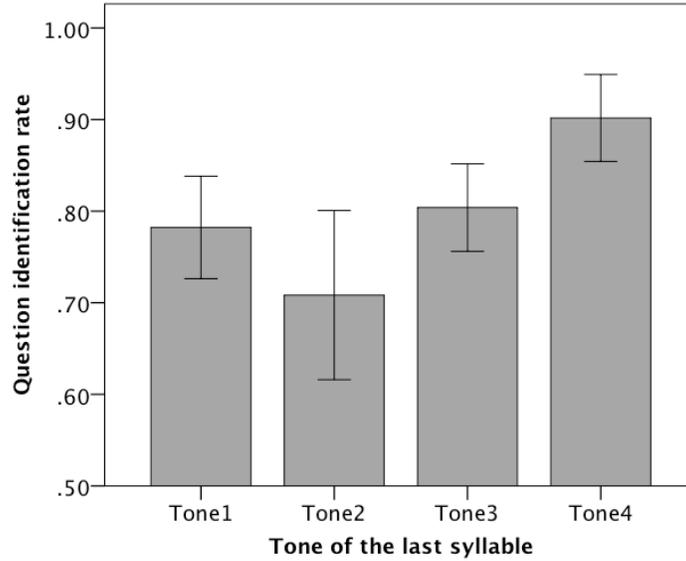
Fig. 2. Mean identification rate of question intonation, grouped by tone of the last syllable. The rate for Tone2 is significantly lower than the other tones, and the rate for Tone4 is significantly higher than the other tones.

**B. Effects of focus on intonation-type identification**

Table 1 compares the intonation identification rates between the utterances with a focus and those without a focus. Different focus positions are each compared with "No focus" using t-tests.

Table I. The t-test results of comparing utterances with and without a focus on intonation-type identification rate.

Intonation type	Focus position (I)	Focus position (J)	Difference (I-J)	p-value
Statement	Beginning	No focus	.006	.684
	Middle	No focus	.004	.763
	End	No focus	-.101(*)	.000
Question	Beginning	No focus	.013	.370
	Middle	No focus	.098(*)	.000
	End	No focus	.097(*)	.000

* The difference is significant at the .05 level.

For statement intonation, as we can see from Table I, only the difference between "End" (a focus at the end, hereafter) and "No focus" was statistically significant. The mean of "End" was smaller than that of "No focus." These results suggest that focus at the beginning or middle position of a sentence does not affect the identification of statement intonation, whereas focus at the end of an utterance makes statement intonation more difficult to identify, as illustrated in Figure 3.
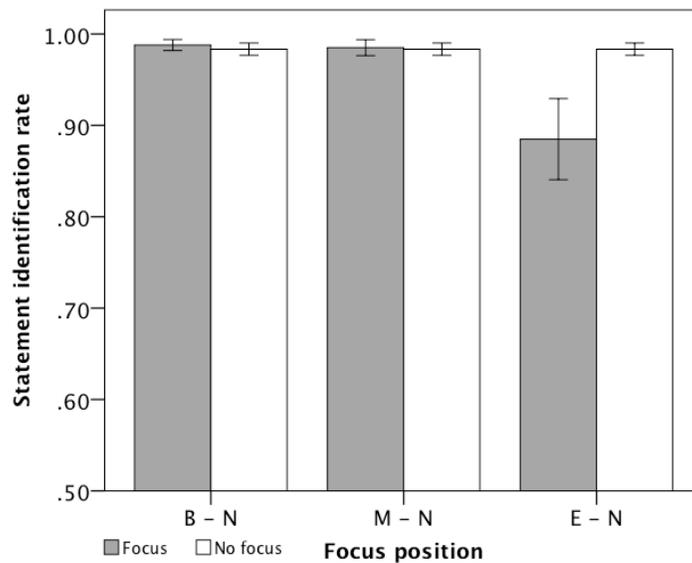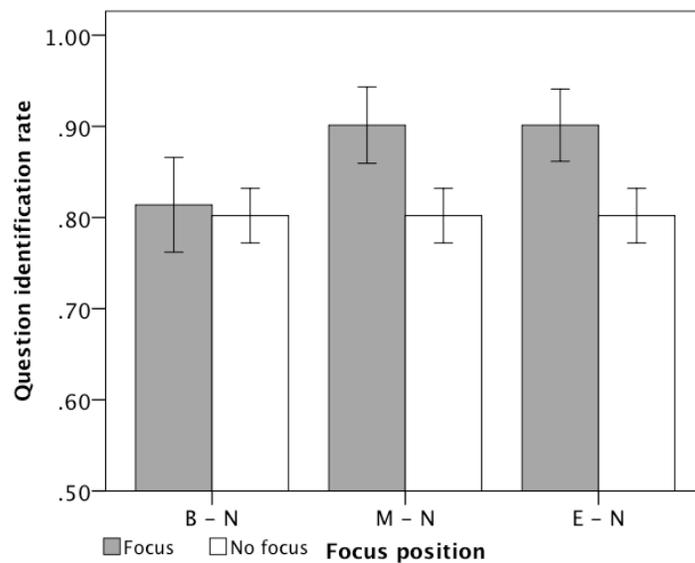

B: Focus at the beginning;   M: Focus at the middle
E: Focus at the end;   N: No focus

Fig. 3. Mean identification rate of statement intonation, comparing different focus positions with "No focus."

We can also see from Table I that for question intonation both the difference between "Middle" and "No focus" and the difference between "End" and "No focus" were statistically significant. The difference between "Beginning" and "No focus," however, was not significant. Additionally, both of the means of "Middle" and "End" were larger than the mean of "No focus." These results suggest that focus at the middle or

the end of a sentence makes question intonation easier to identify whereas focus at the beginning does not affect the identification of question intonation. Figure 4 illustrates the effects of focus at different positions on question intonation identification.

B: Focus at the beginning;   M: Focus at the middle
E: Focus at the end;   N: No focus

Fig. 4. Mean identification rate of question intonation, comparing different focus positions with "No focus."

C. Discriminability and bias

From the signal detection theory point of view (Macmillan and Creelman 2005), to identify statements and questions involves both the ability to discriminate between the two intonation types and the bias in decision-making toward one of the intonation types when the listener is uncertain. Following Pollack and Norman (1964) and Donaldson (1992), we can compute the discriminability and bias measures using formula (1). In this formula, H is the percentage of question intonations that were correctly identified (i.e., hits rate), and FA is the percentage of statement intonations that were incorrectly

identified as a question (i.e., false alarm rate). H and FA were calculated based on the responses from all listeners.

$$A' = 0.5 + [(H-FA)*(1+H-FA)]/[4*H*(1-FA)]$$
$$B'' = [(1-H)*(1-FA)-H*FA]/[(1-H)*(1-FA)+H*FA] \quad (1)$$

A' is a measure of discriminability: scores near 1 indicate high discriminability, while 0.5 indicates chance performance. B'' is a measure of bias ranging from -1 to 1. Positive scores indicate bias toward statement intonation; negative scores indicate bias toward question intonation; and 0 indicates no bias. Table II lists the results, grouped by tone of the last syllable and by focus position, respectively.

Table II. Percentage of questions correctly identified (H), percentage of statements identified as questions (FA), discriminability (A') and bias (B'). The groups based on tone of the last syllable are for the "No focus" utterances only.

Conditions	Groups	H	FA	A'	B''
Tone of the last syllable ("No focus" utterances)	Tone 1	0.782	0.023	0.937	0.841
	Tone 2	0.708	0.021	**0.918**	**0.902**
	Tone 3	0.804	0.003	0.950	0.977
	Tone 4	0.902	0.027	**0.967**	**0.597**
Focus position	No focus	0.802	0.017	**0.944**	**0.872**
	Beginning	0.814	0.012	0.949	0.899
	Middle	0.901	0.015	0.971	0.757
	End	0.901	0.115	**0.940**	**-0.085**

All of the groups except "End" focus had large positive bias scores (B''), suggesting that the perception of intonation has a bias toward statement intonation. When the final tone is Tone2, a rising tone, the bias toward statement intonation (B''= 0.902) was larger than when the final tone is Tone 4, a falling tone (B''=0.597). This result is consistent with the result above that question intonation was harder to identify if the final tone is a rising tone. The Tone4 group also had a higher discriminability score (A' =

0.967) than the Tone2 group (A' = 0.918), suggesting that the two intonation types can be better discriminated for utterances ending with a falling tone than those ending with a rising tone.

Compared to the utterances without a narrow focus, the presence of a focus at the end did not change the discriminability of the two intonation types (A' = 0.944 vs. 0.940). However, it dramatically shifted the bias toward the question intonation (B" = 0.872 vs. -0.085). The discriminability score was highest if there is a focus in the middle of the sentence (A' = 0.971).

IV. DISCUSSION

The tendency across languages to use a rising pitch contour to convey question intonation and a falling contour to convey a statement has prompted many intonation models to present question and statement intonation that are independent of word sequence. While the question intonation tends to be higher than the statement intonation on average, especially near the end of the sentence, intonation contours change not only with social and paralinguistic factors but also with linguistic factors such as tones. In Chinese, question and statement contours are affected by lexical tones to the extent that intonation modeling research using production data fail to extract consistent question and statement contours when sentence materials change (Xu and Wang 1997, Shih 2004). This failure suggests that intonation models need to go beyond simple, tone-independent models in which intonation contours are drawn without considering the tonal composition of the sentence.

This study demonstrates that listeners are not always successful in decoding the multiplexed signal of intonation and lexical tones. The results showed that subjects did not identify statements and questions equally well for sentences with different lexical tone and focus composition. Three asymmetries in the perception of intonation in Mandarin Chinese were revealed from the analyses above: statement and question intonation identification; effects of final focus on statement and question intonation identification; and effects of final Tone2 and Tone4 on question intonation identification.

The asymmetry of statement and question intonation identification is manifested in both identification accuracy and the effect of the last tone. Statement intonation was easier to identify than question intonation; it had both a higher mean identification rate and a smaller variation. The intonation identification experiment was a forced choice test in which the listeners had to identify the intonation type of each utterance as either a statement or a question. Question intonation identification was less accurate, i.e., many question intonation utterances were identified as statements. This result suggests that statement intonation is a default or unmarked intonation type that listeners fall back to when there is not enough information to suggest that the sentence is a question. Question intonation, however, is a marked intonation type. It can be identified only if the listeners actually hear "question" features or mechanisms. The asymmetry of statement and question intonation was also indicated by the finding that the tone of the last syllable did not affect the identification of statement intonation but did affect the identification of question intonation, and by the finding that the listeners had a large bias toward statement intonation. The unequal status of statement and question intonation may not be a language-specific feature of Mandarin Chinese. Makarova (2001), for example, found

that the perception of stimuli as declarative was similar for both Japanese and Russian speakers, while the perception of stimuli as interrogative and exclamatory was significantly different in some cases.

A second asymmetry is that a focus at the end of a sentence made statement intonation more difficult to identify and made question intonation easier to identify. This result is consistent with those of Liu and Xu (2005). From the discriminability and bias analysis above, a focus at the end of a sentence did not change the discriminability of the two intonation types but significantly shifted the bias toward question intonation. Our explanation for this observation is that both question intonation and final focus have a higher prosodic strength at the final position of the sentence (Yuan et al. 2002, Yuan 2004); if there is focus at the end of a statement, the greater strength of the last tone due to the presence of focus may be misinterpreted as a mechanism of question intonation by some listeners, and therefore more statements are identified as questions.

The third asymmetry, the effect of the final tone on question intonation identification, suggests a different scenario. Question intonation was easier to identify for a final falling tone and more difficult to identify for a final rising tone. In contrast to the effect of a final focus on question intonation, a final rising tone was not misinterpreted as a mechanism of question intonation. On the contrary, it caused a greater bias toward statement intonation. This result suggests that tone identification intervenes in the mapping of F_0 contours to intonational categories (i.e., perception of intonation is sensitive to tonal identities). The recognition of tonal identity may precede intonation recognition, or at least tone and intonation are simultaneously recognized. In either case, the phonological identity of the final tone, not just its acoustic form, is involved in the

process of intonation perception. That is, tone and intonation interact at the phonological and linguistic levels. These results demonstrate that intonation in Chinese is not only realized at the level of global or semi-global features, as often claimed in the literature (Fournier et al. 2010, Liang and van Heuven 2007, Vaissiere 2004). Question intonation has a local mechanism that is sensitive to the identity of tones, suggesting that a tone-dependent intonation model is needed. Shih (1988, 2004) put forward the same idea from the acoustic perspective. However, most of the Chinese intonation models in the literature do not capture this aspect of tone and intonation interaction.

V. CONCLUSIONS

The interaction between tone and intonation is of theoretical interest. Our study shows that the interaction between the final tone of a sentence and intonation is present at the phonological level. In contrast, our results do not provide evidence of an interaction between focus and intonation that occurs beyond the psychoacoustic level. Statement intonation is a default, unmarked intonation type, whereas question intonation is a marked type; they do not have equal status. These results help us better understand how the multiplexing of the F_0 channel affects the perception of intonation in Mandarin Chinese, and provide insights that will be useful for further research on the interaction of tone and intonation through both acoustic modeling studies and neurobiological studies.

REFERENCES

Baum, S., and Pell, M. (1999). "The neural bases of prosody: Insights from lesion studies and neuroimaging," *Aphasiology, 13*, 581-608.

Chao, Y. R. (1968). *A Grammar of Spoken Chinese,* Berkeley, CA: University of California Press, pp. 26-39.

Chen, A., Gussenhoven, C., and Rietveld, T. (2004). "Language-specificity in the perception of paralinguistic intonational meaning," *Language and Speech*, 47, 311-349.

Doherty, C. P., West, C., Dilley, L. C., Shattuck-Hufnagel, S., and Caplan, D. (2004). "Question/Statement judgments: an fMRI study of intonation processing," *Human Brain Mapping, 23*, 85-98.

Chen, Y., and Gussenhoven, C. (2008). "Emphasis and tonal implementation in Standard Chinese," *Journal of Phonetics*, 36, 724-746.

Donaldson, D. (1992). "Measuring recognition memory," *J Exp Psychol Gen*, 121, 275-277.

Eady, S. J., and Cooper, W. E. (1986). "Speech intonation and focus location in matched statements and questions," *J. Acoust. Soc. Am*. 80, 402-415.

Fournier, R., Gussenhoven, C., Jensen, O. and Hagoort, P. (2010). "Lateralization of tonal and intonational pitch processing: An MEG study," *Brain Research*, 1328, 79-88.

Fujisaki, H., Wang, C., Ohno, S., and Gu, W. (2005). "Analysis and synthesis of fundamental frequency contours of Standard Chinese using the command-response model," *Speech Communication*, 47, 59-70.

Gandour, J. T. (2007). "Neural substrates underlying the perception of linguistic prosody," In: C. Gussenhoven and T. Riad (Eds.), *Tones and Tunes, Volume 2:*

Experimental studies in word and sentence prosody, Berlin; New York: Mouton de Gruyter, pp. 3-25.

Gandour, J., Tong, Y., Wong, D., Talavage, T., Dzemidzic, M., Xu, Y., Li, X., and Lowe, M. (2004). "Hemispheric roles in the perception of speech prosody," *Neuroimage, 23*, 344-357.

Gårding, E. (1987). "Speech act and tonal pattern in Standard Chinese," *Phonetica*, 44, 13-29.

Gårding, E., Zhang, J., and Svantesson, J. O. (1983). "A generative model for tone and intonation in Standard Chinese based on data from one speaker," *Lund University Working Papers*, 25, 53–65.

Gosy, M., and Terken., J. (1994). "Question Marking in Hungarian: Timing and Height of Pitch Peaks," *Journal of Phonetics*, 22, 269-281.

Grabe, E., Rosner, B., Garcia-Albea, J.E., and Zhou, X. (2003). "Perception of English Intonation by English, Spanish and Chinese Listeners," *Language and Speech*, 46, 375-401.

Gussenhoven, C. (2002). "Intonation and interpretation: Phonetics and Phonology," In *Proceedings of Speech Prosody 2002*, Aix-en-Provence, France, pp. 47-57.

Gussenhoven, C. (2004). *The Phonology of Tone and Intonation*, Cambridge, UK: Cambridge University Press, pp. 71-96.

Gussenhoven, C., and Chen, A. (2000). "Universal and language-specific effects in the perception of question intonation," In *Proceedings of the 6th International Conference on Speech and Language Processing (ICSLP)*, pp. 91-96.

Hadding-Koch, K., and Studdert-Kennedy, M. (1964). "An experimental study of some

intonation contours," *Phonetica*, 11, 175-185.

Ho A. T. (1977). "Intonation Variation in a Mandarin Sentence for Three Expressions: Interrogative, Exclamatory and Declarative," *Phonetica*, 34, 446-457.

House, D. (2003). "Perceiving question intonation: the role of pre-focal pause and delayed focal peak," in *Proceedings of 15th ICPhS*, Barcelona, pp. 755-758.

Jin, S. (1996). *An Acoustic Study of Sentence Stress in Mandarin Chinese*, Ph.D. dissertation, The Ohio State University, Columbus, OH.

Kochanski, G. P., and Shih, C. (2003). "Prosody modeling with soft template", *Speech Communication*, 39, 311-352.

Liang, J., and van Heuven, V. J. (2007). "Chinese tone and intonation perceived by L1 and L2 listeners," In: C. Gussenhoven and T. Riad (Eds.), *Tones and Tunes, Volume 2: Experimental studies in word and sentence prosody*, Berlin; New York: Mouton de Gruyter, pp. 27-61.

Liu, F., and Xu, Y. (2005). "Parallel encoding of focus and interrogative meaning in Mandarin intonation", *Phonetica*, 62, 70-87.

Macmillan, N. A., and Creelman, C. D. (2005). *Detection Theory: A User's Guide (2nd edition)*, Mahwah, New Jersey: Lawrence Erlbaum Associates, pp. 1-512.

Makarova, V. (2001). "Perceptual correlates of sentence-type intonation in Russian and Japanese." *Journal of Phonetics*, 29, 137-154.

Ni, J., and Hirose, K. (2006). "Quantitative and structural modeling of voice fundamental frequency contours of speech in Mandarin," *Speech Communication*, 48, 989-1008.

Ohala, J. (1983). Cross-language use of pitch: an ethological view. *Phonetica*, 40, 1-18.

Peng, S., Chan, M., Tseng, C., Huang, T., Lee, O., and Beckman, M.E. (2005). "Towards

a Pan-Mandarin system for prosodic transcription", In: Sun-Ah Jun (Ed.), *Prosodic Typology: The Phonology of Intonation and Phrasing*, Oxford, UK: Oxford University Press, pp. 230-270.

Pollack, I., and Norman, D. A. (1964). "A non-parametric analysis of recognition experiments," *Psychonomic Science,* 1*,* 125-126.

Prom-on, S., Xu, Y., and Thipakorn, B. (2009). "Modeling tone and intonation in Mandarin and English as a process of target approximation," *J. Acoust. Soc. Am*., 125, 405-424.

Rooth, M. (1985). *Association with Focus*, Ph.D. dissertation, University of Massachusetts, Amherst, MA.

Shen, J. (1992). "Hanyu yudiao moxing chuyi" [On Chinese intonation model], *Yuwen Yanjiu,* 45, 16-24.

Shen, X. (1989). *The Prosody of Mandarin Chinese*, Berkeley: University of California Press, pp. 9-30.

Shih, C. (1988). "Tone and Intonation in Mandarin," *Working Papers of the Cornell Phonetics Laboratory,* 3, 83-109.

Shih, C. (2004). "Tonal effects on intonation," in *Proceedings of International Symposium on Tonal Aspects of Languages: Emphasis on Tone Languages*, Beijing, pp. 163-168.

Studdert-Kennedy, M., and Hadding, K. (1973). "Auditory and linguistic processes in the perception of intonation contours," *Language and Speech*, 16, 293-313.

Tsao, W. (1967). "Question in Chinese," *Journal of Chinese Language Teachers' Association*, 2, 15-26.

Vaissière, J. (2005). "Perception of Intonation," In: D.B. Pisoni and R.E. Remez (Eds.), *The Handbook of Speech Perception*, Malden, MA: Blackwell Publishing, pp. 236-263.

Xu, Y. (1999). "Effects of tone and focus on the formtion and alignment of f_0 contours," *Journal of Phonetics*, 27, 55-105.

Xu, Y., and Wang, Q. E. (1997). "What can tone studies tell us about intonation?" In *Intonation: Theory, Models and Applications, Proceedings of an ESCA Workshop*, Athens, Greece, pp. 337-340.

Xu, Y., and Xu, C. (2005). "Phonetic realization of focus in English declarative intonation," *Journal of Phonetics*, 33, 159-197.

Yuan, J. (2004). *Intonation in Mandarin Chinese: Acoustics, Perception, and Computational Modeling*, Ph.D. Dissertation, Cornell University, NY.

Yuan, J., and Shih, C., (2004). "Confusability of Chinese intonation," In *Proceedings of Speech Prosody 2004*, Nara, Japan, pp. 131-134.

Yuan, J., Shih, C., and Kochanski, G. P. (2002). "Comparison of declarative and interrogative intonation in Chinese," In *Proceedings of Speech Prosody 2002*, Aix-en-Provence, France, pp. 711-714.

Table I. The t-test results of comparing utterances with and without a focus on intonation-type identification rate.

Intonation type	Focus position (I)	Focus position (J)	Difference (I-J)	p-value
Statement	Beginning	No focus	.006	.684
	Middle	No focus	.004	.763
	End	No focus	-.101(*)	.000
Question	Beginning	No focus	.013	.370
	Middle	No focus	.098(*)	.000
	End	No focus	.097(*)	.000

* The difference is significant at the .05 level.

Table II. Percentage of questions correctly identified (H), percentage of statements identified as questions (FA), discriminability (A') and bias (B'). The groups based on tone of the last syllable are for the "No focus" utterances only.

Conditions	Groups	H	FA	A'	B"
Tone of the last syllable	Tone 1	0.782	0.023	0.937	0.841
("No focus" utterances)	Tone 2	0.708	0.021	**0.918**	**0.902**
	Tone 3	0.804	0.003	0.950	0.977
	Tone 4	0.902	0.027	**0.967**	**0.597**
Focus position	No focus	0.802	0.017	**0.944**	**0.872**
	Beginning	0.814	0.012	0.949	0.899
	Middle	0.901	0.015	0.971	0.757
	End	0.901	0.115	**0.940**	**-0.085**

Fig. 1. Mean identification rate of statement intonation, grouped by tone of the last syllable. The differences among the tones are not significant.

Fig. 2. Mean identification rate of question intonation, grouped by tone of the last syllable. The rate for Tone2 is significantly lower than the other tones, and the rate for Tone4 is significantly higher than the other tones.

Fig. 3. Mean identification rate of statement intonation, comparing different focus positions with "No focus."

Fig. 4. Mean identification rate of question intonation, comparing different focus positions with "No focus."