Phonetica 2005;62:70–87
DOI: 10.1159/000090090

Received: June 24, 2005
Accepted: October 4, 2005

Parallel Encoding of Focus and Interrogative Meaning in Mandarin Intonation

Fang Liu[a] Yi Xu[b,c]

[a]Department of Linguistics, University of Chicago, Chicago, Ill., USA;
[b]University College London, Wolfson House, London, UK;
[c]Haskins Laboratories, New Haven, Conn., USA

Abstract

Despite much research, disagreements abound regarding the detailed characteristics of question intonation in different languages or even in the same language. The present study investigates question intonation in Mandarin by also considering the role of focus that is frequently ignored in previous research. In experiment 1, native speakers of Mandarin produced statements, yes/no questions, particle questions, wh-questions, rhetorical questions and confirmation questions with narrow focus on the initial, medial or final word of the sentence, or on none of the words. Detailed F_0 contour analyses showed that focus generated the same pitch range modification in questions as in statements, i.e., expanding the pitch range of the focused word, suppressing (compressing and lowering) that of the post-focus words, but leaving that of the pre-focus words largely unaffected. When the effects of focus (as well as other functions also potentially present) were controlled by subtracting statement F_0 contours from those of the corresponding yes/no questions, the resulting difference curves resembled exponential or even double-exponential functions. Further F_0 analyses also revealed an interaction between focus and interrogative meaning in the form of a boost to the pitch raising by the question starting from the focused word. Finally, subtle differences in the amount of pitch raising were also observed among different types of questions, especially at the sentence-final position. Experiment 2 investigated whether listeners could detect both focus and question in the same utterance. Results showed that listeners could identify both in most cases, indicating that F_0 variations related to the two functions could be simultaneously transmitted. Meanwhile, the lowest identification rates were found for neutral focus in questions and for statements with final focus. In both cases, the confusions seemed to arise from the competing F_0 adjustments by interrogative meaning and focus at the sentence-final position. These findings are consistent with the functional view of intonation, according to which components of intonation are defined and organized by individual communicative functions that are independent of each other but are encoded in parallel.

KARGER

Fax +41 61 306 12 34
E-Mail karger@karger.ch
www.karger.com

Accessible online at:
www.karger.com/journals/pho

Fang Liu
University of Chicago
1010 E. 59th St.
Chicago, IL 60637 (USA)
E-Mail liufang@uchicago.edu

1. Introduction

Every utterance we say in a conversation or a monologue may be of one of several sentence types: statement, question, exclamation, command, or request, among others. In addition to various, often optional, morphosyntactic manipulations, these sentence types are frequently conveyed through prosodic means, pitch contours in particular, or more broadly known as intonation. The difference between statement and question intonation, in particular, has been much researched in many languages. The general consensus is that sentences bearing the meaning of completion, termination, finality or assertion are associated with low or falling pitch, and those bearing the meaning of inquiry, uncertainty, question and non-finality with high or rising pitch [Ladd, 1996]. As summarized by Bolinger [l978], around 70% of the nearly 250 languages examined use a rising terminal to signal questions, whereas others use a higher overall pitch in questions than non-questions. There is much less agreement, however, over the details of such a fall/rise dichotomy.

The first is in regard to the temporal scope of the rise/fall contrast in question versus statement. Many experimental studies have concluded that the relevant acoustic difference only occurs at the end of the sentence; e.g. Chang [1958] for Chengtu Chinese, Fok-Chan [1974], Vance [1976] and Lee [2004] for Cantonese, Rumjancev [1972] and Lin [2004] for Mandarin. Likewise, in the autosegmental and metrical phonology of intonation (AM theory) [Ladd, 1996; Pierrehumbert, 1980], the statement/question contrast is said to be linked only to boundary tones. A boundary tone, transcribed as H% or L% for a high- or low-pitched tone, is defined as a phonological tone located only at the right edge (i.e., the end) of an intonational phrase, although it may take the entire intonational phrase as its association domain.

Studies that have explored longer temporal domains in search of the acoustic correlates of the question/statement contrast have found evidence for nonlocal components. However, the patterns that have been reported are not highly consistent. Two general patterns have been described. The first is that in questions the F_0 of an entire sentence is raised [Haan, 2002; Ho, 1977; Shen, 1990; Yuan et al., 2002]. The other is that the question/statement contrast is time-dependent: the closer to the end of the sentence, the greater the difference between the two sentence types [Lindau, 1986, and Inkelas and Leben, 1990 for Hausa; Ma et al., 2004 for Hong Kong Cantonese; Thorsen, 1978, 1979, 1980 for Danish].

To complicate things further, some researchers have suggested that global F_0 contours of variable shapes are associated with different types of questions. For example, Shen [1990], as also supported by Ni and Kawai [2004] with the same sentence materials, proposes that the feature that distinguishes assertive intonation from interrogative intonation is a difference in register at the starting point: interrogative intonation begins at a higher register than the assertive, but may end with either a high key (in unmarked questions and particle questions) or a low key (in A-not-A questions, alternative questions, and wh-questions).

The complicated temporal patterns reported by Shen [1990] actually suggest that the question/statement contrast should not be investigated independently of other intonational functions. One such function is known as focus, namely, discourse/pragmatic-motivated emphasis. It is now well established that focus plays a critical role in determining the global pitch shape of a declarative sentence. In general, a single (non-final) focus is manifested as tri-zone pitch range adjustments: expanding the pitch range of the focused item, suppressing (lowering and narrowing) the pitch range of all post-focus items, and leaving the pitch range of pre-focus items the same as that in a sentence with no narrow focus [Botinis et al., 2000; Cooper et al., 1985; Selkirk and Shen, 1990; Shen, 1985; Thorsen, 1979; Xu, 1999; Xu and Xu, 2005]. In addition, focus has also been found to be accompanied by an increase in duration on the focused words [Cooper et al., 1985; Xu, 1999]. Furthermore, perception patterns reflexive of such tri-zone pitch range adjustments have been reported [Mixdorff, 2004; Rump and Collier, 1996; Xu et al., 2004]. More importantly, evidence for similar pitch range adjustment has been reported in question intonation as well. For Danish, Thorsen [1980, p. 1021] noted that in sentences containing 'emphasis for contrast', the difference between statement and question 'lies partly in the level and movement of the emphatic syllable, but mainly in the course of the "unstressed" ones after it, which perform less of a fall in questions than in statements.' For English, Eady and Cooper [1986] found that in sentences with initial focus, there is no difference in peak F_0 between the focused word in statements and in questions, but the F_0 toplines depart radically after focus, with statements falling to a low F_0 and questions staying relatively high. For sentences with neutral or final focus, the peak F_0 of the final word in questions is significantly higher than that in statements. For Chinese, Wang [2003] observed that narrow focus is realized in three ways in both statements and questions: the abrupt decrease of the F_0 peak of the syllable following the focused word, the expansion of the pitch range of the focused word and the increase of the F_0 peak of the focused word.

Also as observed by Cooper et al. [1985], Xu and Kim [1996] and Xu [1999], when not given any specific context or instructions, speakers in a recording session often spontaneously emphasize a particular part of a sentence in an unpredictable manner. This means that the occurrence of focus cannot be easily prevented, and thus its effect, if any, cannot be easily avoided. Hence, it is possible that at least some of the discrepancies in the reported question intonation are due to uncontrolled spontaneous focus. In some other cases, the syntactic structure of the sentence may favor a narrow focus on a particular part of the sentence. In Shen's [1990] study, for example, focus can be anywhere in unmarked and particle questions, but in A-not-A questions, focus is likely to occur on the positive component, in disjunctive questions, on the alternative components, and in wh-questions, on the wh-words, especially when used as nouns [see Ishihara, 2002; Li and Thompson, 1979; Tsao, 1967]. Consequently, the phenomena she observed are likely to be the combined effects of interrogative meaning and focus.

It has also been reported that certain additional factors may further affect question intonation, especially its pitch range. Bolinger [1986] has suggested that speaker involvement may affect pitch range: the greater the involvement, the larger the pitch range. Hirschberg and Ward [1992, p. 250] showed that pitch range plays the largest role in interpreting the rise-fall-rise contour (L* + H L H% in ToBI's transcription), with larger pitch ranges indicating incredulity and smaller ones indicating uncertainty. Herman [1996] reported that in Kipare (a Bantu tone language), statements are signalled by nonexpanded pitch range with final lowering, yes/no questions by expanded pitch range with final lowering, and incredulous questions by expanded pitch range with final raising. Jun and Oh [1996] suggested that for some Korean speakers, incredulity questions (echo questions expressing incredulity) are distinguished from wh-questions by a larger pitch range, higher amplitude, and boundary tones.

The above discussion shows that several issues still need to be resolved about question intonation. First, it is not yet clear whether question intonation involves F_0 variations only at the sentence-final position, or rather in a larger temporal domain. The final-only hypothesis is essential to the notion of boundary tone in the AM theory of intonation. Although studies such as that by Eady and Cooper [1986] have shown that there are F_0 differences nonlocal to the final word between statements and questions, it has been argued that the nonlocal patterns can all be accounted for in terms of phonetic implementation of sequential phonological units, involving L% boundary tone plus downstep for statement, but H% plus suspension of downstep for question [Ladd, 1996]. Since downstep is assumed in the AM theory as a phonetic implementation rule triggered only by certain pitch accents such as H*L [Pierrehumbert and Beckman, 1988], it is possible to resolve this issue in a language like Mandarin where sentences can be found consisting of only H tones, thus preventing downstep from being triggered. Second, the role of focus in shaping question intonation is not yet fully clear: does focus involve the same tri-zone pitch range adjustments in question as in statement? Third, the conflicting findings about whether the F_0 of an entire question is raised need to be resolved. In this respect, we note the frequent mention in the literature of the meanings nonessential to the interrogative meaning of questions, such as incredulity or surprise, and the possible link between these nonessential meanings and global pitch raising. Thus, there is a need to separate these meanings from the interrogative meaning when investigating question intonation.

The present study was therefore designed to address three issues regarding question intonations in Mandarin. (1) Does the question/statement contrast involve pitch differences only at the sentence-final position or over a larger temporal domain? (2) Can focus and interrogative meaning be produced and perceived together in question intonation? If yes, do they also interfere with each other? (3) Is it possible to separate the phonetic manifestation of interrogative meaning from those of noninterrogative meanings, e.g. incredulity, confirmation, and rhetoric? Two experiments were conducted to answer these questions. Experiment 1 examined the acoustic patterns related to focus and interrogative meaning in several sentence types. Experiment 2 tested whether focus and interrogative meaning could be perceived simultaneously by Mandarin listeners.

2. Experiment 1

The goal of experiment 1 is to investigate the acoustic manifestations of question intonations in Mandarin by addressing the following questions. (1) How does focus interact with interrogative meaning in determining the F_0 contours of questions? (2) What are the basic constituents of question intonation? (3) Are there F_0 differences among different types of questions (yes/no question, particle question, wh-question, rhetorical question, and confirmation question) in Mandarin?

2.1. Methods

2.1.1. Materials

Four basic sentence frames (each consisting of 10 syllables, all having identical tones: high, rising, low or falling, corresponding to tone 1, 2, 3 or 4) were used, as shown in table 1. These sentence frames were converted to six sentence types (statement, yes/no question, particle question, wh-question, rhetorical question, and confirmation question) by alternately adding an interrogative pronoun or verb phrase [*shúi* ('who/whom'), *gànmá* ('do what')], a negative particle [*bùshì* ('not')], a yes/no particle [*shìbùshì* ('yes/no')], an interrogative particle *(ma)*, a period, and/or a question mark. The sentences were to be said with focus at four possible locations (initial, medial, final, and none, i.e., neutral focus). Seventy-six

Table 1. Basic sentence frames used for constructing test materials

	Focus/key word				
	initial		medial		final
Frame 1 (tone 1, high)	*ZhāngWēi*	dānxīn	*XiāoYīng*	kāichē	*fāyūn*
	ZhangWei	worry	*XiaoYing*	driving	*dizzy*
	'*ZhangWei* worries that *XiaoYing* will get *dizzy* while driving'				
Frame 2 (tone 2, rising)	*WángMéi*	huáiyí	*LiúNíng*	huáchuán	*zháomí*
	WangMei	suspect	*LiuNing*	canoeing	*obsessed*
	'*WangMei suspects* that *LiuNing* will get *obsessed* with canoeing'				
Frame 3 (tone 3, low)	*LǐMǐn*	fǎngǎn	*LiǔYǔ*	diǎnhuǒ	*qǔnuǎn*
	LiMin	dislike	*Liu Yu*	light a fire	*keep warm*
	'*LiMin* dislike *Liu Yu* to light a fire to *keep warm*'				
Frame 4 (tone 4, falling)	*YèLiàng*	hàipà	*ZhàoLì*	shùijiào	*zuòmèng*
	YeLiang	afraid	*ZhaoLi*	sleep	*dream*
	'*YeLiang* is afraid that *ZhaoLi* will *dream* while sleeping'				

distinct sentences were constructed by varying tone component, sentence type, and focus location. Each sentence was to be repeated 5 times by each subject. Therefore, a total of 3,040 sentences (76 sentences × 5 repetitions × 8 subjects) were investigated. The F_0 contours of three key words in each sentence, shown as italicized in table 1, were extracted and measured.

The following is the set of sentence types composed from frame 1 with all high tones.

Statement:
ZhāngWēi dānxīn *XiāoYīng* kāichē *fāyūn*.
'*ZhangWei* worries that *XiaoYing* will get *dizzy* while driving.'

Yes/no Question:
ZhāngWēi dānxīn *XiāoYīng* kāichē *fāyūn*?
'*ZhangWei* worries that *XiaoYing* will get *dizzy* while driving?'

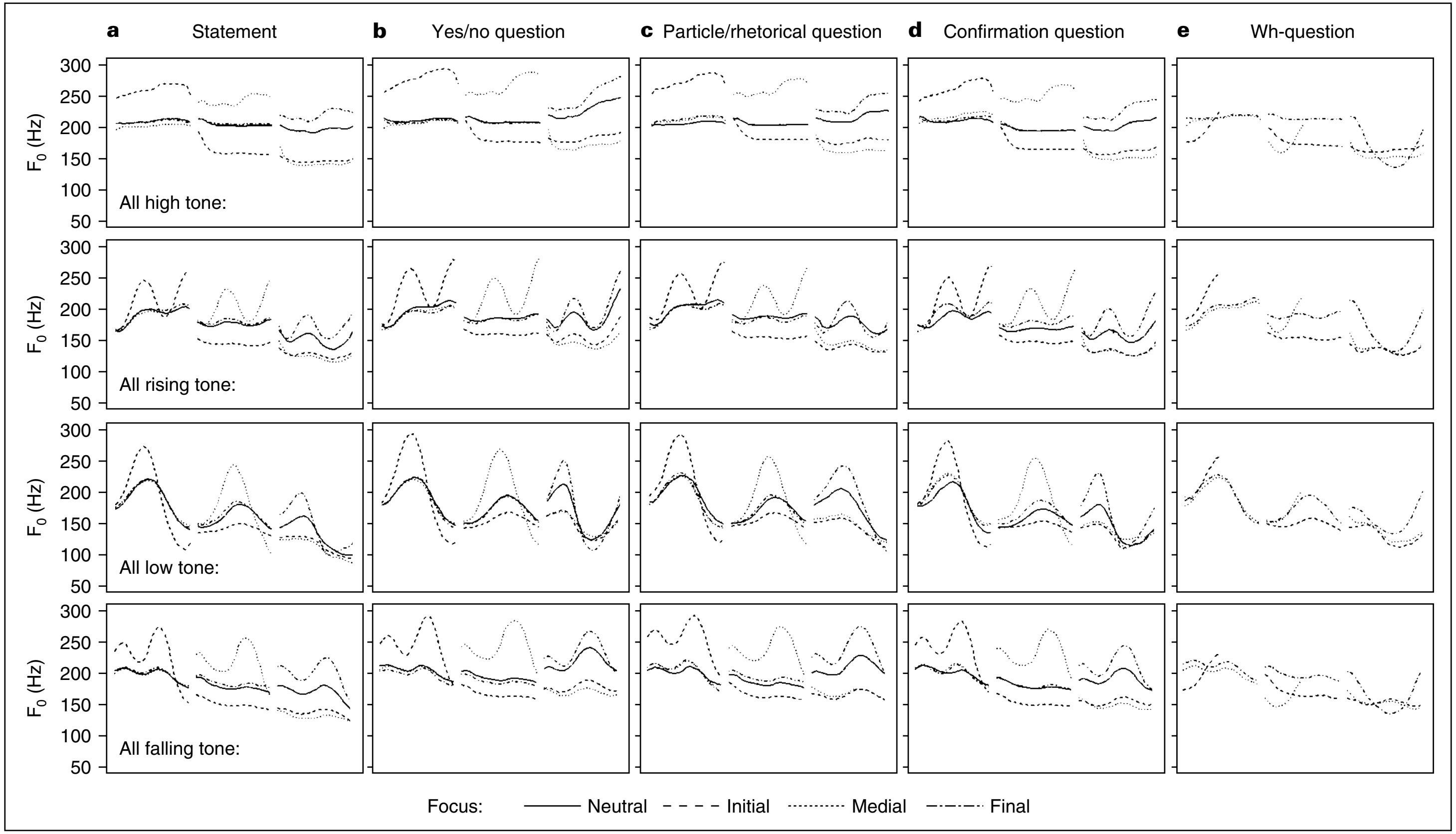
Wh-Question:
Shúi dānxīn *XiāoYīng* kāichē *fāyūn*?
'*Who* worries that *XiaoYing* will get *dizzy* while driving?'

Particle Question:
ZhāngWēi dānxīn *XiāoYīng* kāichē *fāyūn* ma?
'Does *ZhangWei* worry that *XiaoYing* will get *dizzy* while driving?'

Rhetorical Question:
Bùshì *ZhāngWēi* dānxīn *XiāoYīng* kāichē *fāyūn* ma?
'Isn't it *ZhangWei* who worries that *XiaoYing* will get *dizzy* while driving?'

Confirmation Question:
Shìbùshì *ZhāngWēi* dānxīn *XiāoYīng* kāichē *fāyūn*?
'Is it the case that *ZhangWei* worries that *XiaoYing* will get *dizzy* while driving?'

Fig. 1 The effect of focus on global F_0 of different sentence types in the four sentence frames. In each graph, the curves separated by the breaks are the F_0 contours of the initial, medial and final key words, averaged across all the repetitions and individual speakers. All the curves are time-normalized. The F_0 shapes of the wh-words are very different from those of other words at the same position because they have different syllabic and tonal compositions (see section 2.1.1). In the all low tone sentences, the first low tone in each word is changed into rising tone due to a phonological rule [Chao, 1968].


a Statement
b Yes/no question
c Particle/rhetorical question
d Confirmation question
e Wh-question
F_0 (Hz)
300
250
200
150
100
50
All high tone:
All rising tone:
All low tone:
All falling tone:
Focus: Neutral Initial Medial Final

2.1.2. Subjects

Eight native speakers of Mandarin, 4 males and 4 females, served as subjects. They were either students at Yale University or residents in New Haven, Conn., USA, born and raised in the city of Beijing where Mandarin is the vernacular. They had no self-reported speech and hearing disorders and their ages ranged from 22 to 34.

2.1.3. Recording

Recording was done in a sound-treated booth at Haskins Laboratories, New Haven, Conn. A JavaScript program running under a web browser controlled the flow of the recording. The subject was seated comfortably in front of a computer screen, wearing a headset microphone. The microphone was about 2 inches away from the left side of the subject's lips. The target sentences were displayed on a computer screen, one at a time, in random order. Subjects were instructed to say each sentence as a statement or question depending on whether it ended with a period or a question mark, and to emphasize any word that was surrounded by square brackets. The utterances were directly digitized onto a hard disk at 44.1 kHz sampling rate and 16-bit amplitude resolution. The digitized sound was later resampled at 22.05 kHz.

2.1.4. F_0 Extraction and Measurement

Using a custom-written script for the Praat program (www.praat.org), the waveform and spectrogram of each sentence and a label window were displayed automatically on a computer monitor. Onset and offset labels were manually inserted for the three key words of the sentence whose F_0 trajectories and duration measurements were to be taken. Vocal pulse markings generated by Praat were displayed in another window. The pulses were inspected and any erroneous markings (such as missing pulses and double markings) were manually corrected. A Perl program then read in all the segment files and F_0 files saved by the Praat script. It applied a trimming algorithm to remove local spikes in the F_0 curves [Xu, 1999].

The Perl program then extracted various measurements, including the mean, maximum, and minimum F_0 values of target words and their locations. Each mean F_0 value was computed by averaging over all the F_0 points in a word. For visual inspection and graphic analysis, the Perl program also computed time-normalized F_0 contours by getting the same number of evenly spaced F_0 points from each keyword. Durations of the key words were also taken by the program. This allowed the display of average F_0 contours against average time, assuring minimal information loss.

2.2. Results

For direct visual comparison, average F_0 curves (from 40 repetitions by 8 subjects) of the three key words with different lexical tones, under different focus conditions, and in different sentence types are displayed in figures 1 and 2. In computing these curves, the F_0 values were converted to a logarithmic scale before averaging, so as not to bias the means toward speakers with a larger F_0 range. The mean values were converted back to hertz after averaging.

2.2.1. Effect of Focus on the Global F_0 Curve

Figure 1 shows the average F_0 contours of the three key words in different sentence types (statement, yes/no question, particle question, rhetorical question, confirmation question, and wh-question) and focus (neutral, initial, medial, and final) with the all high, rising, low, and falling tone sentence frames, respectively. What can clearly be seen is that regardless of sentence type and lexical tone, the pitch range of the focused words is raised and expanded (expansion most apparent in the low tone), that of the post-focused words compressed and lowered, and that of the pre-focused words largely unaffected.

Repeated-measures ANOVAs on duration and mean F_0 of initial, medial, and final key words were conducted, with lexical tone (high, rising, low, and falling), focus (neutral, initial, medial, and final) and sentence type (statement, yes/no question, and

Table 2. Results of repeated-measures ANOVAs on the effect of focus on the global F_0 curve

Key words	Duration, ms	Mean F_0, Hz
Initial	$F(3, 21) = 35.501$, $p < 0.0001$ Initial (418.33) > final (283.99), medial (279.59), neutral (278.72)	$F(3, 21) = 41.579$, $p < 0.0001$ Initial (249.80) > neutral (203.40), final (201.59), medial (199.67)
Medial	$F(3, 21) = 49.14$, $p < 0.0001$ Medial (378.00) > final (241.09), neutral (230.87), initial (225.62)	$F(3, 21) = 73.284$, $p < 0.0001$ Medial (231.37) > final (188.68), neutral (188.01) > initial (159.76)
Final	$F(3, 21) = 26.828$, $p < 0.0001$ Final (408.97) > neutral (348.64), initial (348.62), medial (345.11)	$F(3, 21) = 50.282$, $p < 0.0001$ Final (211.69) > neutral (189.24) > initial (149.52), medial (148.10)

confirmation question) as fixed factors and subjects as replication factor (table 2). Both duration and mean F_0 of focused words are found to be increased, regardless of sentence type and tone composition. Furthermore, mean F_0 of the final key words under neutral focus is only slightly lower than that under final focus (though marginally significant according to a linear mixed-effects regression model: $t = -2.13$, $p = 0.0338$). The suppression effect of focus on post-focused words is manifested by the significantly lowered mean F_0 of the medial key words under initial focus and by the significantly lowered mean F_0 of the final key words under initial and medial focus.

2.2.2. *Effect of Sentence Type on the Global F_0 Curve*

Figure 2 shows mean F_0 contours of the three key words in different sentence types and tone compositions, and under different focus conditions. As can be seen, in sentences with initial or medial focus, the difference between question and statement is manifested as a moderate raise in pitch range, starting from the focused word. Focus thus serves as a pivot at which statement and question contours start to diverge. In sentences with final or neutral focus, the difference between statement and question is manifested mainly in the final word, suggesting that the widely recognized question intonation with a final rise is that of a question with final or neutral focus.

To compare the effects of sentence type at different sentence locations, repeated-measures ANOVAs on mean F_0 of unfocused initial, medial, and final key words were conducted, with lexical tone (high, rising, low, and falling), focus (initial, medial, and final, but no neutral focus because both wh- and rhetorical questions have an implicit narrow focus) and sentence type (statement, yes/no question, wh-question, rhetorical question, and confirmation question) as fixed factors and subjects as replication factor. As can be seen, the differences in mean F_0 among the sentence types increase as the sentence approaches the final position (see also figure 3). Thus, the greatest difference among sentence types is found at the sentence-final position. Pitch raising by question intonation is greater in yes/no and rhetorical questions than in confirmation and wh-questions (table 3). Intonation of particle questions and wh-questions is different from that of statement, indicating that in addition to the use of a particle and wh-word, pitch raising also occurs in these questions.

2.2.3. *Interaction of Focus and Sentence Type*

Repeated-measures ANOVA on post-focus pitch drop [= F_0 (st) of focused key word – F_0 (st) of post-focus key word] was conducted, with lexical tone (high, rising,

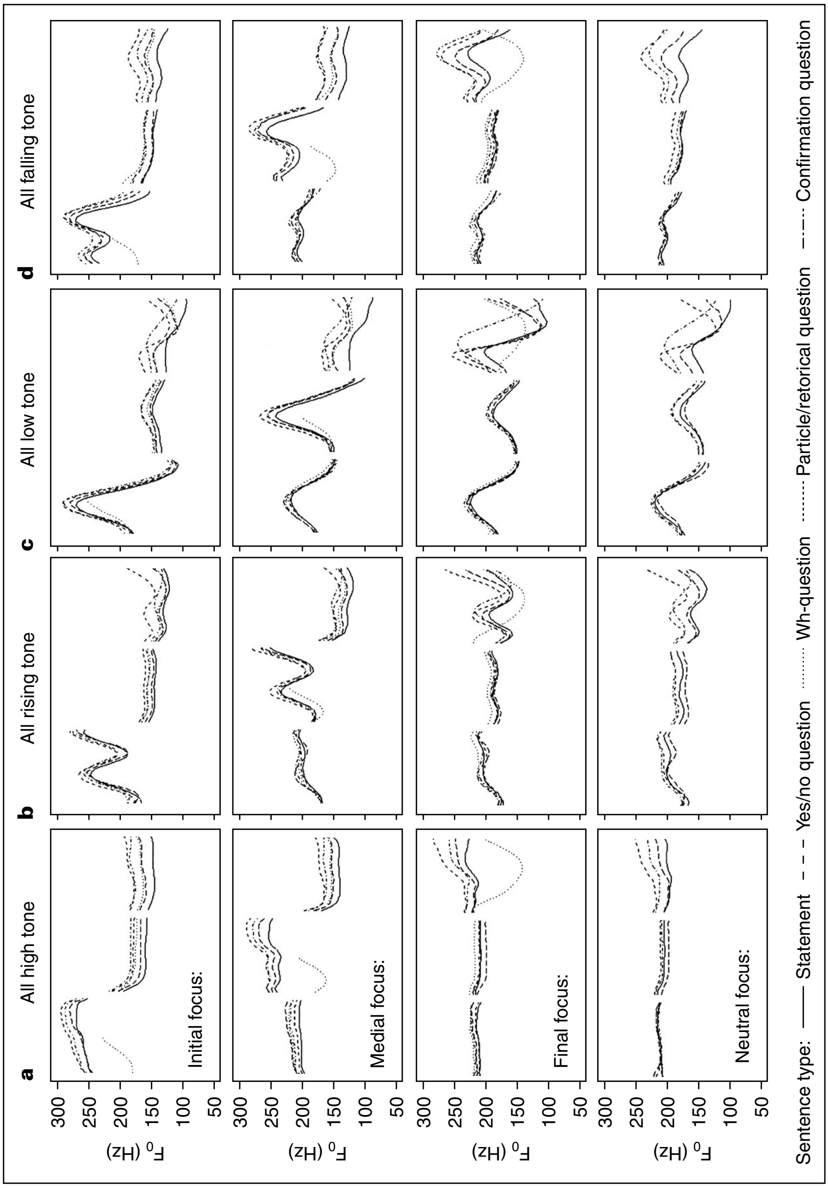


Fig. 2 The effect of sentence type on global F_0 of the four sentence frames in different focus conditions. See caption of figure 1 for detailed explanations.

low, and falling), focus (initial and medial) and sentence type (statement, yes/no question, rhetorical question, and confirmation question) as fixed factors and subjects as replication factor. Post-focus pitch drop is not significantly different between initial and medial focus conditions ($F < 1$). Thus, initial and medial focus have similar amounts of suppression effects on the post-focus words. Questions with stronger

Table 3. Results of repeated-measures ANOVAs on the effect of sentence type on the global F_0 curve

Key words	Mean F_0, Hz
Initial	$F(4, 28) = 1.51, p = 0.2261$ Wh-question (204.77), confirmation question (203.09), rhetorical question (202.98), yes/no question (200.26), statement (198.54)
Medial	$F(4, 28) = 5.848, p = 0.0015$ Rhetorical question (181.52), yes/no question (179.71), wh-question (179.41) > confirmation question (172.57), statement (170.38)
Final	$F(4, 28) = 21.817, p < 0.0001$ Yes/no question (163.26), rhetorical question (161.74) > wh-question (152.13), confirmation question (148.51) > statement (134.67)

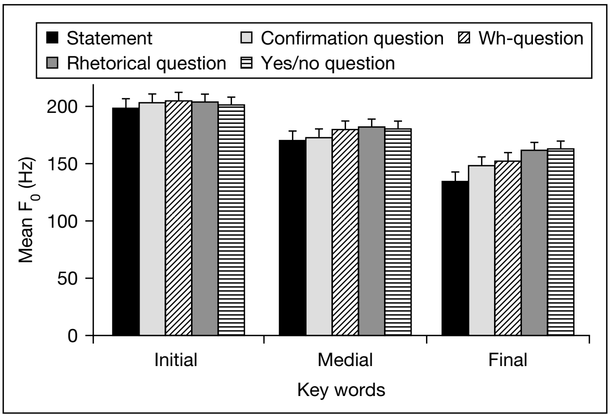


Fig. 3 Mean F_0 of the key words across the entire sentences under different sentence types.

incredulity connotations slightly reduce the degree of suppression of post-focus words [$F(3, 21) = 3.239$, $p = 0.0427$], as indicated by the following ordering of post-focus pitch drop: statement (8.55), confirmation question (8.22) > yes/no question (7.55), rhetorical question (7.19).

2.2.4. Nonlinearity of Question-Statement Difference

The analyses in section 2.2.2 show that the $\mathbf{F_0}$ raising in a question accelerates toward the end of the sentence. It is therefore possible that the raising is non-linear. To examine this possibility, difference $\mathbf{F_0}$ curves were obtained by subtracting the $\mathbf{F_0}$ value of statements from that of yes/no questions. In the difference curves, confounding factors common to both question and statement, new topic in particular [Xu, 2005], are largely left out, leaving only the 'pure' contrast between the two sentence types. Figure 4 displays the difference curves averaged across repetitions, tones (excluding the low tone) and subjects and grouped by focus. The columns group the fitted curves and equations obtained through three types of regressions: linear, exponential (both using mean time as predictor) and double-exponential (using time squared as predictor). The dependent variables are the corresponding **mean $\mathbf{F_0}$ values in semitones plus 1** (to make all values

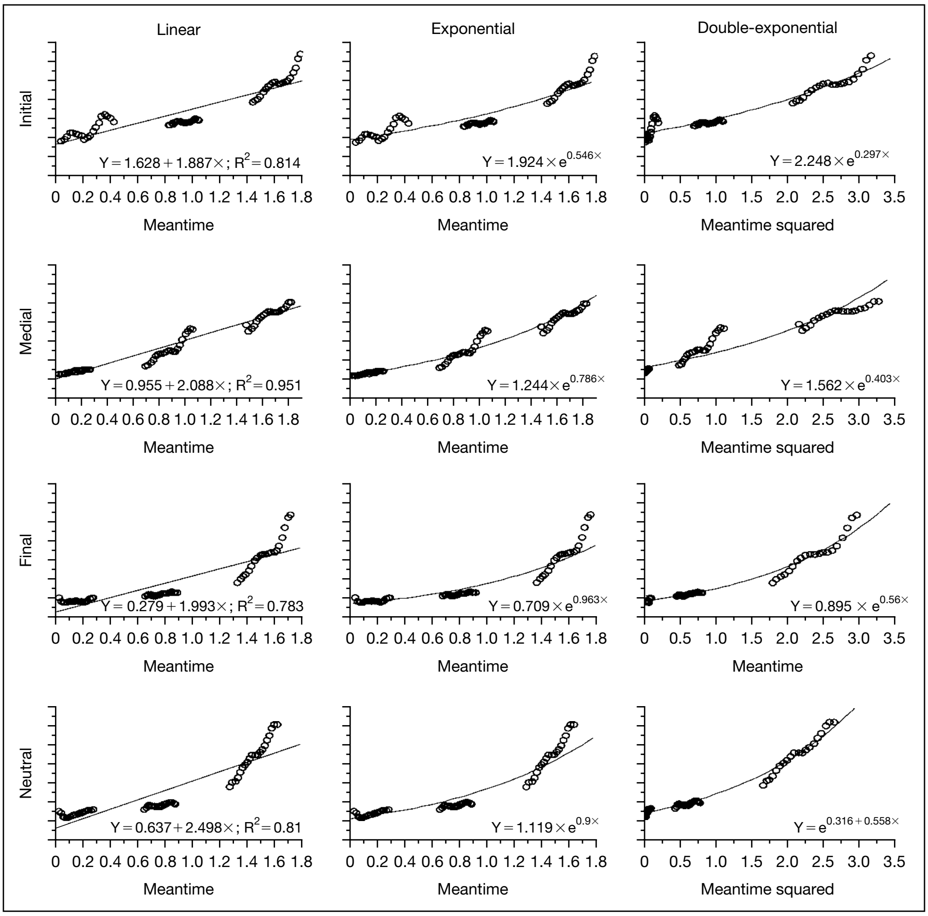


Fig. 4 Circles = Mean difference F_0 in semitones (F_0 of yes/no question – statement +1) averaged over the four basic sentence frames and grouped by focus conditions (initial, medial, final, neutral). Thin curves = Fitted curves obtained through linear, exponential and equivalent of double-exponential regressions. Equations of the curves are at the bottom of each graph.

Table 4. *t* values of linear, exponential and double-exponential regressions for four focus types

Focus type	Regression type		
	linear	exponential	double-exponential
Neutral	15.57	24.12	54.79
Initial	15.79	18.75	23.42
Medial	33.36	41.37	21.75
Final	14.35	25.56	55.61

positive so that the exponential regressions are operable). As can be seen, except for medial focus, the curve fitting gets better from linear to exponential and to double-exponential. Table 4 shows the *t* values of the three types of regressions grouped by focus type (*p* values are all less than 0.0001 and thus not displayed). With the only exception of medial focus, the *t* values are always the largest in the double-exponential regression and the second largest in the exponential regression. Thus, the 'pure' difference between question and statement does not appear to be linear, but at least exponential, or even double-exponential.

In figure 4, one can also see the effect of focus on the statement/question difference. That is, the difference receives a boost at the location of the focus. This boost has probably led to the much smaller *t* values of the double-exponential regressions for initial and medial focus than for neutral and final focus as seen in table 4.

2.3. Discussion

The results of the above analyses suggest that the effects of focus are present in both statements and questions: the pitch range of the focused words is expanded, that of the post-focus words compressed and lowered, and that of the pre-focus words largely unaffected. The whole pitch level is shifted upward in questions with initial focus (in comparison with the corresponding statements). In sentences with medial focus, the difference between questions and statements is manifested as a moderate raise in pitch range starting from the focused words. Focus thus serves as a pivot at which statement and question contours start to diverge. $\mathbf{F_0}$ of both statements and questions with no narrow focus (neutral focus) is similar to those with final focus, i.e., showing the greatest sentence type difference in the final syllable. This seems to be evidence that the widely recognized question intonation with an extensive final rise is that of a question with final or neutral focus. The gradual pitch range raise caused by interrogative meaning is greater in yes/no questions and rhetorical questions than in wh-questions and confirmation questions, suggesting a possible separation of incredulity and interrogation as independent functions.

3. Experiment 2

Experiment 2 was designed to address the second question raised in the introduction, namely, whether native listeners are able to simultaneously perceive the statement/question contrast and the presence and location of focus.

3.1. Methods

3.1.1. Materials

Three hundred and twenty statements and yes/no questions by two of the speakers (1 male, 1 female; 2 speakers × 4 tones × 4 foci × 2 sentence types × 5 repetitions = 320) in experiment 1 were used as stimuli.

3.1.2. Subjects

Eleven native speakers of Mandarin, 5 males and 6 females, served as subjects in this experiment. They were either students at Yale University or residents in New Haven, Conn., born and raised in the city of Beijing. They had no self-reported speech and hearing disorders and their ages ranged from 22 to 34. Five of them also served as speakers in experiment 1.

3.1.3. Procedure

The task of the subject was to identify the sentence type (statement or question) and focus (initial, medial, final, or none, explained to them as emphasis[1]) of each of the 320 sentences. During the test, the subject was seated comfortably in front of a computer screen in a quiet room, wearing a headphone set. In each trial, eight response categories (question/initial, question/medial, question/final, question/none, statement/initial, statement/medial, statement/final, statement/none) were displayed as selection boxes on the computer screen, and the subject clicked on the one that matched his/her impression after hearing each sentence. A new sentence was played 1.0 s after a choice was made. The whole process took about 45 min on average.

3.2. Results

There were 3,520 trials (11 subjects × 320 trials) in total, of which 89.12% and 88.72% were perceived with correct sentence type and focus, respectively. Among the 713 misperceived trials (20.26%), 67 were perceived with both wrong sentence type and wrong focus (1.90%), 316 with only wrong sentence type (8.98%), and 330 with only wrong focus (9.38%).

Table 5 is a matrix showing the percentage of classification of each of the categories (row) to all eight categories (column). The following mismatching patterns can be observed. (1) Statement and question with initial or medial focus are most likely confused with each other. This is attributable to the effect of post-focus suppression: initial and medial focus compress and lower the pitch range of the post-focused words, making the pitch range of the final word in a question close to that of the final word in a statement. (2) Statements with final focus are the least easy to recognize (12.50% were identified as statements with neutral focus, 13.86% as questions with final focus, and 7.73% as questions with neutral focus). This indicates the similarity between neutral and final focus for both statement and question, and highlights the competing effects of sentence type and final focus on the F_0 contour of the final word: when the pitch range of the final word in a statement is raised by focus and thus somewhat resembling the F_0 pattern of the final word in a question, it is likely for the two sentence types to be confused. (3) The second most difficult to recognize are questions with neutral focus, of which 22.73% were heard as questions with final focus, and 6.14% as statements with neutral focus. Again, this is due to the similarity in F_0 between neutral and final focus. (4) 10.91% of the questions with final focus are recognized as questions with neutral focus. Once more, this is due to the similarity in F_0 between neutral and final focus.

Table 6 displays the mean accuracy rate of focus perception for each lexical tone grouped by sentence type and focus, collapsed across gender and focus location. Repeated-measures ANOVAs on focus perception were conducted, with speaker (male and female), focus (initial, medial, final, and neutral), lexical tone (high, rising, low, and falling), and sentence type (statement and yes/no question) as fixed factors and listener as replication factor. The perception of the female speaker has a significantly higher accuracy rate (90.6%) than that of the male speaker (86.9%) [$F(1, 10) = 5.343$, $p = 0.0434$]. The accuracy rates of perception for different focus locations are significantly different

[1] It is critical to instruct the subjects to listen for emphasis as opposed to listening for prominence as done in some studies. The identification of the former is quite robust as demonstrated by a number of studies [Mixdorff, 2004; Rump and Collier, 1996; Xu et al., 2004].

Table 5. Matrix of classification percentage (%) for each combination of sentence type and focus

	S_Initial	S_Medial	S_Final	S_Neutral	Q_Initial	Q_Medial	Q_Final	Q_Neutral
S_Initial	**89.32**	0.23	0	3.18	6.82	0	0	0.45
S_Medial	0.23	**86.82**	0.23	2.05	0	10.45	0.23	0
S_Final	0.68	0.45	**63.86**	12.50	0.68	0.23	13.86	7.73
S_Neutral	5.00	1.14	4.32	**81.14**	0.45	0.23	0.23	7.50
Q_Initial	12.27	0	0.23	0.91	**82.27**	0	2.72	1.59
Q_Medial	0	12.50	0.23	0.23	0.45	**84.55**	0.91	1.14
Q_Final	0	0	2.27	0.91	0.23	0.45	**85.23**	10.91
Q_Neutral	0.45	0.23	1.82	6.14	2.95	0.91	22.73	**64.77**

S = statement; Q = question.

Table 6. Mean accuracy rate of focus perception for each lexical tone grouped by sentence type and focus (collapsed across gender)

Sentence type	Focus	Lexical tone				Mean
		high	rising	low	falling	
Statement	initial	95.5	97.3	98.2	93.6	96.2
	medial	98.2	97.3	97.3	96.4	97.3
	final	85.5	70.9	70	84.5	77.7
	neutral	90	90.9	87.3	86.4	88.7
Question	initial	93.6	97.3	94.5	92.7	94.5
	medial	95.5	97.3	96.4	99.1	97.1
	final	81.8	86.4	95.5	86.4	87.5
	neutral	74.5	68.2	74.5	66.4	70.9
Mean		89.3	88.2	89.2	88.2	88.7

[$F(3, 30) = 12.37$, $p < 0.0001$]. A Student-Newman-Keuls post hoc test indicates that accuracy rates of medial (97.2%) and initial (95.3%) focus are significantly higher than those of final (82.6%) and neutral (79.8%) focus. Lexical tone has no significant effects on the perception of focus. However, the interactions of sentence type × focus [$F(3, 30) = 31.28$, $p < 0.0001$], sentence type × tone [$F(3, 30) = 3.37$, $p = 0.0313$], and sentence type × focus × tone [$F(9, 90) = 3.60$, $p = 0.0007$] are all significant. This is possibly due to the following perception results. (1) Final focus in statements and neutral focus in questions are the hardest to identify. (2) Focus perception in questions with high and falling tones is worse than that in questions with rising and low tones, which is exactly opposite to the trend of focus perception in statements.

Table 7 displays the mean accuracy rates of sentence type perception, collapsed across gender and focus location. Repeated-measures ANOVAs on sentence type perception were conducted, with speaker (male/female), focus (initial, medial, final, and neutral), lexical tone (high, rising, low, and falling), and sentence type (statement and yes/no question) as fixed factors and listeners as replication factor. The perception of the female speaker (93.4%) is more accurate than that of the male speaker (84.8%) [$F(1, 10) = 80.91$, $p < 0.0001$]. The accuracy rates for statements and questions are not significantly different. Both focus [$F(3, 30) = 3.02$, $p = 0.045$; neutral (91.5%) > initial

Table 7. Mean accuracy rates of sentence type perception for each lexical tone grouped by sentence type and focus (collapsed across gender)

Sentence type	Focus	Lexical tone				Mean
		high	rising	low	falling	
Statement	initial	92.7	91.8	92.7	93.6	92.7
	medial	85.5	84.5	92.7	94.5	89.3
	final	62.7	73.6	78.2	95.5	77.5
	neutral	89.1	92.7	87.3	97.3	91.6
Question	initial	68.2	87.3	96.4	94.5	86.6
	medial	71.8	85.5	99.1	91.8	87.1
	final	95.5	95.5	98.2	98.2	96.9
	neutral	84.5	90	95.5	95.5	91.4
Mean		81.3	87.6	92.5	95.1	89.1

(89.7%) > medial (88.2%) > final (87.2%)] and lexical tone [$F(3, 30) = 48.06$, $p < 0.0001$; falling (95.1%) > low (92.5%) > rising (87.6%) > high (81.2%)] have significant effects on the perception of sentence type. Student-Newman-Keuls post hoc tests indicate that the perception of sentence type under neutral focus is significantly better than under final focus, and that the pairwise differences between lexical tones for the perception of sentence type are all statistically significant. The interactions of sentence type × focus [$F(3, 30) = 25.48$, $p < 0.0001$], sentence type × tone [$F(3, 30) = 4.36$, $p = 0.0116$], focus × tone [$F(9, 90) = 3.91$, $p = 0.0003$], and sentence type × focus × tone [$F(9, 90) = 6.74$, $p < 0.0001$] are all significant, which is reflexive of the following facts. (1) The identification of statement under final focus has the lowest accuracy rate among the four focus locations, and initial and medial focuses cause trouble for the identification of question. (2) Questions with high tone are hard to identify. (3) Sentences with high tone have the lowest accuracy rates across all four focus conditions.

3.4. Discussion

Overall, the results of experiment 2 show that listeners could identify both sentence type (statement versus yes/no question) and focus most of the time (89.1% and 88.7%, respectively). Nevertheless, low accuracy rates were found for neutral focus in questions (71%) and statements with final focus (78%). These confusions seemed to arise from the competing F_0 adjustments by sentence type and focus at the sentence-final position (final F_0 was raised for both question and final focus, but not for statement).

4. General Discussion and Conclusion

The results of the present study largely answered two of the three questions raised in the introduction. First, a clearer picture than before has emerged regarding the general pattern of question intonation. Previously, question intonation has been described as involving (a) boundary tone only [Pierrehumbert, 1980; Ladd, 1996; Lin, 2004], (b) raising of F_0 of the entire sentence [Haan, 2002; Ho, 1977; Shen, 1990; Yuan et al., 2002], or (c) superposition of a linear baseline onto the sentence starting from the first accented word [Thorsen, 1980]. The current results are most consistent with account c, but the detailed acoustic data also reveal that the global function is not linear, but more likely exponential or even double-exponential. Such exponential raising, unlike the linear raising proposed by Thorsen, can explain the much larger F_0 raising at the end of a question, which has been the major motivation for the boundary tone account. But the exponential shape also demonstrates that the accelerated final rise is only part of the global question function, albeit the most prominent phase of the function.

Second, both the production and perception data show that focus and interrogative meaning can be transmitted concurrently. In production, focus exerts similar effects on the overall F_0 contours of questions as on those of statements, i.e., expanding the on-focus pitch range, suppressing the post-focus pitch range, and leaving the pre-focus pitch range largely neutral. Interrogative meaning raises pitch over the course of the sentence in an accelerated manner resembling an exponential or double-exponential function. In perception, listeners can simultaneously perceive both focus and interrogative meaning with high accuracy in most cases.

The current data have also provided preliminary evidence for answering the third question raised at the outset of the study. That is, it is possible to separate the phonetic manifestation of interrogative meaning from that of noninterrogative meanings. As seen in table 3, the pitch range raise by question intonation is greater in yes/no and rhetorical questions than in confirmation and wh-questions. As shown by Hu [2002], speakers tend to raise the pitch register of the entire sentence to express surprise. Ho [1977] has found that F_0 is even higher in an exclamatory than in an interrogative sentence. The order of pitch raising found in the present study seems to agree with the amount of incredulity/surprise in the different types of questions. However, because sentences in the present study were produced without context or realistic discourse interaction, only subtle differences were observed across the question types. Clearer separations may be made in future studies using paradigms that involve realistic context or dialogue interactions to control the element of incredulity/surprise.

Regarding the complex global F_0 pattern related to different types of questions as reported by Shen [1990], the current data also provide possible explanations. As seen in the difference curves in figure 4, in addition to the exponential raising, F_0 in a question receives an extra boost at the location of the focus. Thus, if focus consistently occurs in certain sentence structures, such as in A-not-A questions, alternative questions, and wh-questions, the extra boost would interact with the global F_0 raising by question, increasing the on-focus pitch as well as decreasing the post-focus pitch. As for why F_0 is boosted at focus in a question, again it is possibly related to the element of incredulity/surprise that seems to be closely related to focus. That is, while focus in a statement serves to highlight the importance of the focused item, in a question it seems to naturally carry a connotation of surprise/incredulity: is it really *that thing/person* that you mean?

The current data do not imply, however, that the encoding scheme of interrogative meaning in a language can only involve exponential pitch raising found in the present study. It is highly likely that there exist language-specific components related to question intonation. In English, for example, in addition to the final pitch raising, a nonfinal focused word in a question probably has low, rather than high pitch, as in a statement [Eady and Cooper, 1986]. This is apparently not the case in the preset data, presumably because in a tone language the local underlying pitch targets are not easily changed, for they encode lexical information. In Greek, Hungarian, Romanian and Neapolitan Italian, it is known that F_0 always drops at the end of a yes/no question rather than rises as in English [D'Imperio, 2002; Grice et al., 2000]. It is possible that in these languages the local pitch targets are also not free to change, just as in the case of Mandarin falling tone, in which F_0 drops even in a question. But also like Mandarin, questions in these languages may involve nonlocal pitch raising, as is already reported for Neapolitan Italian [D'Imperio, 2001].

Last but not least, the new data appear to have provided a key to solving the long-standing puzzle as to why final focus is manifested much less effectively than an earlier focus [Botinis and Bannert, 1997; Botinis et al., 1999; Botinis et al., 2000; Cooper et al., 1985; Jin, 1996; Rump and Collier, 1996; Xu, 1999]. Because the sentence-final position is where the exponential pitch raising generates the greatest distinction between statements and questions, pitch range modification by any other function at the end of a sentence would directly compete with the question intonation. As shown by the results of experiment 2, final focus in a statement, which already raises F_0 much less than does an earlier focus, still led listeners to often hear the sentence as a question rather than a statement with final focus. It thus seems that it is this competition that has been preventing final focus from significantly raising pitch range at the end of a sentence.

Overall, the findings of the present study seem to support the functional view of intonation, according to which components of intonation are defined and organized by individual communicative functions that are independent of each other [Xu, 2005]. These functions are encoded in parallel, each with an encoding scheme that is distinct from those of all other functions. There are nevertheless frequent interactions among the encoding schemes because of limited availability of acoustic/articulatory dimensions and space, which has resulted in a delicate balance between functions that share the same articulatory/acoustic parameters. However, each communicative function has to have at least one dominant encoding characteristic for it to be functional; the dominance of that encoding characteristic would lead to compromises by other functions that sometimes also have a need to use a similar encoding characteristic, as is the case with final focus versus question.

Acknowledgements

We owe our thanks to Alan Yu and Gina Levow for very helpful comments on an earlier version of the paper. This work is supported in part by NIH Grant R01DC03902 and NIH grant 1R01DC006243 to the second author.

References

Bolinger, D.: Intonation across languages; in Greenberg, Universals of human language, vol. 2, Phonology, pp. 471–523 (Stanford University Press, Stanford 1978).

Bolinger, D.: Intonation and its parts: melody in spoken English (Stanford University Press, Palo Alto 1986).

Botinis, A.; Bannert, R.: Tonal perception of focus in Greek and Swedish. Proc. ESCA Workshop on Intonation: Theory, Models and Applications, Athens 1997, pp. 47–50.

Botinis, A.; Bannert, R.; Tatham, M.: Contrastive tonal analysis of focus perception in Greek and Swedish; in Botinis, Intonation: analysis, modelling and technology, pp. 97–116 (Kluwer Academic Publishers, Boston 2000).

Botinis, A.; Fourakis, M.; Gawronska, B.: Focus identification in English, Greek and Swedish. Proc. 14th Int. Congr. Phonet. Sci., San Francisco 1999, pp. 1557–1560.

Chang, N.T.: Tones and intonation in the Chengtu dialect (Szechuan, China). Phonetica *2:* 59–85 (1958).

Chao, Y.R.: A grammar of spoken Chinese (University of California Press, Berkeley 1968).

Cooper, W.E.; Eady, S.J.; Mueller, P.R.: Acoustical aspects of contrastive stress in question-answer contexts. J. acoust. Soc. Am. *77:* 2142–2156 (1985).

Eady, S.J.; Cooper, W.E.: Speech intonation and focus location in matched statements and questions. J. acoust. Soc. Am. *80:* 402–416 (1986).

Fok-Chan, Y.Y.: A perceptual study of tones in Cantonese (University of Hong Kong Press, Hong Kong 1974).

Haan, J.: *Speaking of questions: an exploration of Dutch question intonation*; PhD diss. Nijmegen (2002).

Herman, R.: Final lowering in Kipare. Phonology *13:* 171–196 (1996).

Hirschberg, J.; Ward, G.: The influence of pitch range, duration, amplitude and spectral features on the interpretation of the rise-fall-rise intonation contour in English. J. Phonet. *20:* 241–251 (1992).

Ho, A.T.: Intonation variation in a Mandarin sentence for three expressions: interrogative, exclamatory and declarative. Phonetica *34:* 446–457 (1977).

Hu, F.: A prosodic analysis of wh-words in Standard Chinese. Proc. 1st Int. Conf. Speech Prosody, Aix-en-Provence 2002, pp. 403–406.

Inkelas, S.; Leben, W.R.: Where phonology and phonetics intersect: the case of Hausa intonation; in Kingston, Beckman, Papers in laboratory phonology 1 – Between the grammar and physics of speech, pp. 17–34 (Cambridge University Press, Cambridge 1990).

Ishihara, S.: Syntax-phonology interface of wh-constructions in Japanese. Proc. 3rd Tokyo Conf. Psycholinguist., Tokyo 2002, pp. 165–189.

Jin, S.: *An acoustic study of sentence stress in Mandarin Chinese*; PhD diss. Ohio State University, Columbus, Ohio (1996).

Jun, S.-A.; Oh, M.: A prosodic analysis of three types of wh-phrases in Korean. Lang. Speech *39:* 37–61 (1996).

Ladd, D.R.: Intonational phonology (Cambridge University Press, Cambridge 1996).

Lee, W.-S.: The effect of intonation on the citation tones in Cantonese. Proc. Int. Symp. Tonal Aspects Lang. Emphasis Tone Lang., Beijing 2004, pp. 107–110.

Li, C.N.; Thompson, S.A.: The pragmatics of two types of yes-no questions in Mandarin and its universal implications. 15th Reg. Meet. Chicago Linguist. Soc., Chicago 1979, pp. 197–206.

Lin, M.: On production and perception of boundary tone in Chinese intonation. Proc. Int. Symp. Tonal Aspects Lang. Emphasis Tone Lang., Beijing 2004, pp. 125–129.

Lindau, M.: Testing a model of intonation in a tone language. J. acoust. Soc. Am. *80:* 757–764 (1986).

Ma, J.K.-Y.; Ciocca, V.; Whitehill, T.L.: The effects of intonation patterns on lexical tone production in Cantonese. Proc. Int. Symp. Tonal Aspects Lang. Emphasis Tone Lang., Beijing 2004, pp. 133–136.

Mixdorff, H.: Quantitative tone and intonation modeling across languages. Proc. Int. Symp. Tonal Aspects Lang. Emphasis Tone Lang., Beijing 2004, pp. 137–142.

Ni, J.-F.; Kawai, H.: Pitch targets anchor Chinese tone and intonation patterns. Proc. Int. Conf. Speech Prosody, Nara 2004, pp. 95–98.

Pierrehumbert, J.: *The phonology and phonetics of English intonation*; *PhD* diss. MIT, Cambridge, Mass. (1980).

Pierrehumbert, J.; Beckman, M.: Japanese tone structure (MIT Press, Cambridge 1988).

Rumjancev, M.K.: Ton i intonacija v sovremennom kitajskom jazyke (Izdatel'stvo Moskovskogo Universiteta, Moscow 1972).

Rump, H.H.; Collier, R.: Focus conditions and the prominence of pitch-accented syllables. Lang. Speech *39:* 1–17 (1996).

Selkirk, E.; Shen, T.: Prosodic domains in Shanghai Chinese; in Inkelas, Zec, The phonology-syntax connection, pp. 313–337 (University of Chicago Press, Chicago 1990).

Shen, J.: Beijinghua shengdiao de yinyu he yudiao; in Lin, Wang, Beijing Yuyin Shiyanlu, pp. 73–130 (Beijing University Press, Beijing 1985).

Shen, X.S.: The prosody of Mandarin Chinese (University of California Press, Berkeley 1990).

Thorsen, N.: An acoustical investigation of Danish intonation. J. Phonet. *6:* 151–175 (1978).

Thorsen, N.: Lexical stress, emphasis for contrast, and sentence intonation in advanced standard Copenhagen Danish. Proc. 9th Int. Congr. Phonet. Sci., Copenhagen 1979, pp. 417–423.

Thorsen, N.G.: A study of the perception of sentence intonation – Evidence from Danish. J. acoust. Soc. Am. *67:* 1014–1030 (1980).

Tsao, W.-Y.: Question in Chinese. J. Chinese Lang. Teach. Assoc. *2:* 15–26 (1967).

Vance, T.J.: An experimental investigation of tone and intonation. Phonetica *33:* 368–392 (1976).

Wang, A.: *Putonghua Yudiao Yingao Jiangshi Yanjiu*; PhD diss. Beijing (2003).

Xu, Y.: Effects of tone and focus on the formation and alignment of F_0 contours. J. Phonet. *27:* 55–105 (1999).

Xu, Y.: Speech melody as articulatorily implemented communicative functions. Speech Commun. *46:* 220–251 (2005).

Xu, Y.; Kim, J.: Downstep, regressive upstep, H-raising, or what? – Sorting out the phonetic mechanisms of a set of 'phonological' phenomena. J. acoust. Soc. Am. *100:* 2824 (1996).

Xu, Y.; Xu, C.X.: Phonetic realization of focus in English declarative intonation. J. Phonet. *33:* 159–197 (2005).

Xu, Y.; Xu, C.X.; Sun, X.: On the temporal domain of focus. Proc. Int. Conf. Speech Prosody, Nara 2004, pp. 81–84.

Yuan, J.; Shih, C.; Kochanski, G.P.: Comparison of declarative and interrogative intonation in Chinese. Proc. 1st Int. Conf. Speech Prosody, Aix-en-Provence 2002, pp. 711–714.