



How does Prosody Distinguish Wh-statement from Wh-question? A Case Study of Standard Chinese

Xuefei Liu, Aijun Li, Yuan Jia

Institute of Linguistics, Chinese Academy of Social Sciences, Beijing, China

liuxuefei1234@163.com, liaj@cass.org.cn, summeryuan_2003@126.com

Abstract

There are wh-sentences which express speech acts of interrogative or declarative with the same syntactic structure in standard Chinese, such as, "bǎobao chīdiǎnr shénme?" (What does the baby intend to eat?) and "bǎobao chīdiǎnr shénme." (The baby intends something to eat.) The interrogative pronoun "shénme" (what) has different semantic functions, such as specific reference in the interrogative sentence, and indefinite reference in the declarative sentence. The current paper focuses on the prosodic aspect of these kinds of wh-sentences based on some well-designed dialogues. Prosodic features are divided into local and global ones. Local features include prosodic cues of boundary syllables, the potential prominent word, the wh-words and the difference of prosody features between wh-words and the following boundary syllable. The global features include F0 spans, the regression of lines and the first order difference of F0 of the whole sentence. Fisher discrimination analysis shows that both global and local prosodic features contribute to discriminate the speech acts of those wh-sentences, but local features are more reliable than global features. The results indicate that features representing intonation components, such as sentence stress, boundary tones or even prosodic structures, must be considered in speech act discrimination besides syntax and context.

Index Terms: speech act, prosodic features, wh-question, Chinese intonation, specific reference, indefinite reference

1. Introduction

Recognition of two major speech acts, i.e. statement and interrogative, is a big challenge in dialogue understanding and speech-to-speech translation. However, it is still a problem unsolved, especially in Chinese. As we all know that the interrogative pronoun is the typical syntactic marker for wh-questions. However, in spoken Chinese the sentences with interrogative pronoun, whether the speech act is statement or interrogative cannot be decided by the syntax of the sentence on its own, because Chinese does not need a wh-movement like English. Instead, the interrogative pronouns stay "in-situ", that is, where the answer to the question will go. Accordingly, the sentence like "bǎobao xiǎng chīdiǎnr shénme" should have two speech acts "What does the baby intend to eat?" or "The baby intends something to eat." The interrogative pronouns (wh-words) of Standard Chinese can act as a specific reference in the interrogative act, or an indefinite reference meaning, such as "anything", "anyone", "something" or "someone" as in the declarative act.

In this case, the speech acts should only be discriminated by prosodic or contextual information. Therefore, in this paper

we will introduce a pilot study on these kinds of wh-sentences which have the same syntactic structure but different speech acts, in order to check how effective the prosody to discriminate the two speech acts, and which prosodic features are more reliable with the discrimination.

1.1. Functional Distributions of Wh-words

The functional distribution of wh-words were examined in different speaking styles in previous studies [1,2]. The result indicated that the wh-words acting as interrogatives account for 50~60% in the scenarios of news reports, interviews or online chats, but the ratio raised to about 80% in call center services on taxi, airport or restaurant, etc. This implies that the research on the speech act of wh-sentences is significant in speech interaction system.

Moreover, the speech act of some wh-sentences can be decided by the grammatical forms alone, such as the grammatical form "wh-words +dōu" is generally a wh-declarative with arbitrary reference meaning of wh-word, here is an example "shénme shì hǎo dōuxíng." (Anytime is ok.)[2].

1.2. Sketch of the Intonation of Standard Chinese

The ripple-wave relation between Chinese tones and intonation was first investigated by Chao [3,4]. Lin and Li [5] proposed a feature-based model of Chinese intonation which consists of two key components, focal prominence and boundary tones. Chinese intonation, like non-tonal languages, can also be adequately described in the framework of intonational phonology [6]. They summarized the features as below:

Narrow Focus: [+RAISE H] and [+LOWER L]
Interrogative boundary tone: [+RAISE TONE]
Declarative boundary tone: [+LOWERTONE]

Liu and Xu [7] proposed a 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.

Following these views, we summarize the research on Chinese intonation on interrogative and statement by intonational components. Unfortunately, we have not found a unique component or feature that has the power to discriminate the two intonations.

For the stress of wh-sentences, there is a dispute on the stress location relating to syntactic constituents. Some claimed that although interrogative pronoun is the default focus of the interrogative sentence, sentential stress may be placed on other constituents [8-9]; others believed that interrogative pronoun is also the constituent to bear the sentence stress [10-12]. However, opinions are consistent on wh-declarative, because

wh-word is not the focus of the sentence, it has few chance to acquire the sentential stress accordingly [4].

For intonational question, final boundary tone plays an important role in distinguishing statements from interrogatives [7, 13-15]. But for questions like yes-no questions and wh-questions, the final boundary tone is not a reliable signal to discriminate it from declarative sentences [16-19].

Besides stress and boundary tone, acoustic analysis showed that the global intonation patterns differ among statements and interrogatives [14,20-22], where subtle differences in the amount of pitch raising were observed among different types of questions, especially at the sentence final position [14-15]. The ‘pure’ difference between question and statement does not appear to be linear, but at least exponential, or even double-exponential [23, 24].

1.3. Sentence Type Classification or Modeling

Lots of studies had been conducted to classify or model the Chinese intonational questions and statements. Liu et al. [23] used decision trees with three different sets of feature vectors to determine the most significant elements in an utterance that signify the sentence type. They confirmed the previous findings that the difference between statement and question intonations in Mandarin is manifested by an increasing departure from a common starting point toward the end of the sentence. They also found that the parameters corresponding to the sentence-final F0 are found to be the dominant factor of determining the sentence type. Nevertheless, F0 precede the final syllable is also consistently found to be relevant.

However, most of other studies on sentence type classification only consider the overall quantification of the intonation trend, such as in [19], only the regression line of F0 trajectory is simulated to distinguish unmarked questions from statements.

In the current study, we follow the idea proposed by Liu and Xu [7], Liu et al. [23] and Liu [24], and divide the prosodic features into local and global ones according to the intonation components. The local features are related to the prosodic components of boundary tones, the wh-words and the difference between wh-words and the following noun phrases. The global features are the prosodic features of the whole sentence.

The present research focuses on the above mentioned wh-sentences which have the same syntactic structures but different speech acts of interrogative or declarative. Two issues are concerned: 1) The intonation patterns to express the two speech acts; 2) How effective are the prosodic features to differentiate these two speech acts?

2. Speech Data

2.1. Dialogue Materials

The target sentence adopted here is a 9-syllable sentence with the frame: “xiǎozhāng dāsuan V.diǎnr shénme NP”, where wh-word is shénme (what). Both V and NP were designed as monosyllabic words in 4 tonal contexts: H tone T1, rising tone T2, low dipping tone T3 and high falling tone T4. Speech acts of interrogative and declarative are realized in two dialogue contexts.

1) Declarative context where wh-word “shénme” has indefinite reference meaning, the target statement is in broad focus.

A: fāshēngle shénmeshǐ? (what's wrong?)

B:xiǎozhāng dāsuan chī(ná/mǎi/mài)diǎnr shénme gāo (yú/mǐ/cài).

XiaoZhang intends to eat(bring/buy/sell) whatever cake(fish/rice/dish).

2) Interrogative context where wh-word “shénme” has specific reference meaning:

A: xiǎozhāng dāsuan chī(ná/mǎi/mài)diǎnr shénme gāo(yú/mǐ/cài)?

What kind of cake(fish/rice/dish) does XiaoZhang intend to eat (bring/buy/sell)?

B: zǎogāo (dǎyú/xiǎomǐ/bōcài).

Cake (fish/rice/dish).

So the number of prompts for recording is 2 speech acts*4 V tones*4 NPs tones=32 dialogues.

2.2. Speakers and Recording

8 speakers of standard Chinese (4 males and 4 females, average age is 24.13) were recruited to read the 32 dialogues in a random order in a sound proof room. All speakers were born in Beijing and had no hearing problem. They were allowed to repeat the recording if they made mistakes. Finally, 32*8=256 target sentences were obtained for the acoustic analysis.

2.3. Annotation

Both segmental and prosodic annotations were made with PRAAT (<http://www.fon.hum.uva.nl/praat/>) including segmental boundaries of syllables (initial/final), prosodic boundaries and stress levels. F0 data were automatically extracted and manually checked by the authors. For each syllable, F0 data were extracted and interpolated into 10-points.

2.4. Data Normalization of F0

All the prosodic data were extracted in Praat. F0 data were normalized through Z-Score method in order to eliminate speakers' differences. For each individual speaker, the z-scored F0 was calculated with formula (1).

$$F0z = (F0 - F0_{i,mean}) / F0_{i,\sigma} \quad (1)$$

Where $F0_{i,mean}$ is the mean F0 of speaker i , $F0_{i,\sigma}$ is the standard deviation of F0 of speaker i . F0 is in semitone scale referred by speaker i 's bottom f0 in hertz.

3. Analysis and Results

3.1. Intonation Patterns

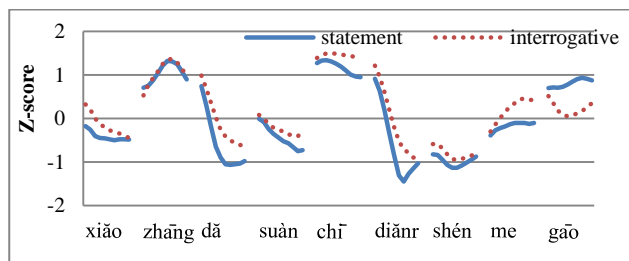


Figure 1: Mean F0 contours of wh-statements and wh-interrogatives for T1(V, chī)+T1(N, gāo) structure with duration normalized for each syllable. (statement: XiaoZhang intends to eat whatever cake./ interrogative: What cake does XiaoZhang intend to eat?)

The mean F0 contours of statements and interrogatives for T1+T1 combinations are plotted in Figure 1. It shows that the global F0 trajectory of wh-interrogatives is higher than that of statements except the pitch of boundary tone. According to the nonparametric analysis of 2 related samples, the maximum and the range of the first order difference in these two different moods show significant difference, while other parameters show no difference. This means that the discrimination of wh-sentences should investigate local features besides the global ones.

About the boundary tones, final boundary tones show a high tonal feature "H%" for wh-interrogative as seen in Figure 1. The H tone of wh-declarative boundary is mainly caused by the stressed final NP rather than the intonation. When comparing the tonal patterns of final boundary tones, we find that tone 1(H level tone) of wh-interrogative becomes a rising pattern, while it keeps an H level tone in wh-declaratives. The rising terminal should be the signal of an interrogative decoded by the boundary tone.

For the sentence stress, according to the prosodic annotation, the sentence stress tends to be located at the final boundary syllables, specifically, 91.4% for wh-statements, 73.44% for wh-questions; but stress on the verb is 8.59% for wh-statements and 9.38% for wh-questions; while stress on wh-words is 17.19% for wh-questions and 0% for wh-statements.

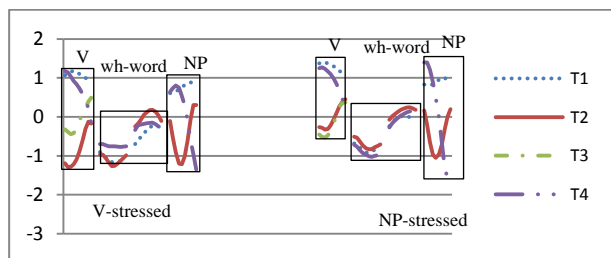


Figure 2: Mean F0 contours of "V+wh-words+NP" in wh-statements with various tones of V and NP in 2 stress conditions

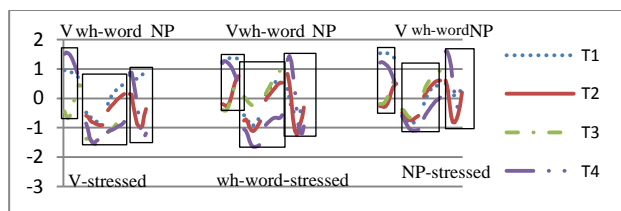


Figure 3: Mean F0 trajectories of "V+wh-words+NP" in wh-questions with various tones of V and NP in 3 stress conditions

Figures 2 and 3 depict the F0 trajectories of the final three words in wh-statements and wh-questions with various tones of "V" and "NP" in different stress conditions, respectively. The boxes show the pitch ranges of the words. The figures show that stressed words have either larger pitch range or higher register than those of unstressed words, and there is PFC (post focus compression) in both two speech acts. However there are salient differences between the two speech acts in the following aspects:

(A) Wh-word

The pitch range of Wh-word is significantly smaller in wh-statements than that in wh-questions without regarding the distribution of sentence stress. And the wh-word in wh-statements never attracts sentential stress, which indicates that if the sentence stress locates in the wh-words, the sentence can be recognized as wh-question.

(B) Boundary tone

In wh-statements the boundary tones keep the tonal shapes as the lexical tones (see Fig. 2), but the tonal shapes change a lot in wh-questions (see Fig. 3), especially the final part of the tone. T1 and T4 become a rising pattern, and T2 and T3 change the rising slope, which are encoded by the boundary feature H% of the interrogative intonation. Besides the slope of the boundary tone, other acoustic features relating to the interrogative boundary tone are manifested by a higher bottom value.

(C) The F0 difference between wh-words and the boundary tone

The F0 difference between wh-words and the boundary tones is much larger in wh-statements than that in wh-questions irrespective of the intonational stress.

3.2. Discriminate Analysis

Based on the phonetic analysis on intonations of the two speech acts, we propose local features in addition to globe features in the discriminate analysis.

3.2.1. Feature Extraction

Global and local features are extracted to apply to the discriminate analysis based on the analysis of intonation patterns of wh-sentences.

1) Local features

- features of boundary syllables (Bt)

F0:

Bt-F0_{max}, Bt-F0_{min}, Bt-F0_{mean}, Bt-F0_{range};

Slope: Bt-F0_{slope}

First order difference (FD) of F0:

Bt-FD_{max}, Bt-FD_{min}, Bt-FD_{mean}, Bt-FD_{range}, Bt-FD_{slope}

Duration:

Duration of boundary syllables including duration of initials (Bt-smddur); duration of finals (Bt-ymdur); duration of syllables (Bt-dur).

Intensity:

Intensity of boundary syllables (Bt-Int)

HNR:

Harmonicity to Noise Ratio of each finals (HNR)

- features on wh-word "shénme"(what)

F0 and duration of wh-word: wh-F0_{max}, wh-F0_{min},

wh-F0_{mean}, wh-F0_{range}, wh-dur and wh-HNR

Difference of F0 (duration/HNR) between F0 of wh-

word and the following boundary syllable(NP): D-F0_{max},

D-F0_{min}, D-F0_{mean}, D-F0_{range}, D-ymdur, D-dur, D-Int, D-

HNR, D-FD_{max}, D-FD_{min}, D-FD_{mean}, D-FD_{range}.

2) Global features

Global features include F0 spans and regression lines of

whole sentences: S-F0_{max}, S-F0_{min}, S-F0_{mean}, S-F0_{range}, S-LR.

First order difference of F0 of the whole sentences: S-FD_{max},

S-FD_{min}, S-FD_{mean}, S-FD_{range}, S-FD_{LR}.

3.2.2. Discriminate Analysis

Fisher discriminate analysis (stepwise method) was adopted to investigate the significance of the prosodic features to discriminate these two speech acts. All the extracted data were grouped by boundary tones according to the "boundary tone" theory of Lin [5]. The analysis was carried out in 3 ways: global features only, local features only and "global+local" merging features. The discrimination results are listed in Table 1.

Table 1. Discrimination accuracy in 3 ways

boundary tones	global	local	global+local
T1	62.50%	92.20%	95.30%
T2	70.30%	87.50%	87.50%
T3	62.50%	90.60%	90.60%
T4	68.80%	93.80%	95.30%

The results in table 1 illustrate that local features play more important role than global features. Merging local and globe features obtained the best results, which is up to 95% in some situations (T1 and T4), and local features produced rather close results.

3.3. Discriminant and Classification Coefficients of Individual Prosodic Feature

From the analysis, we could also obtain the linear regression and the discriminant coefficients for each prosodic feature as shown in Table 2.

Table 2. Linear regression coefficient and the discriminant coefficient of the prosodic features

Boundary tones	parameters	wh-statement	wh-question	Coefficients
T1	wh-HNR	1.078	1.76	0.935
	D-	-2.411	2.388	0.881
	S-LR	62.406	323.408	0.468
	wh-	-6.586	43.29	0.334
	D-vmdu	-111.737	-161.139	-0.444
	(Constant)	-14.915	-28.689	--
T2	Bt-slope	-2.939	-20.466	0.703
	D-HNR	-0.101	-0.337	0.575
	D-F0max	-10.775	-13.478	0.465
	D-F0min	1.908	-0.339	0.382
	wh-HNR	1.163	1.628	-1.068
	(Constant)	-28.961	-34.24	--
T3	Bt-dur	165.49	129.056	-0.625
	D-HNR	0.429	0.054	-0.796
	wh-HNR	1.138	1.632	0.771
	wh-	-47.016	174.841	0.58
	S-max	4.808	8.085	0.523
	(Constant)	-53.766	-51.842	--
T4	Bt-vmdu	63.855	116.464	0.514
	D-HNR	-0.018	-0.336	-0.726
	wh-HNR	0.53	1.16	1.061
	S-	-175.43	-90.105	0.624
	wh-	-144.287	29.003	0.424
	(Constant)	-38.758	-42.088	--

The statistics in the last column represent the discriminant coefficients for each boundary tone type of wh-sentences, which stand for the effect of the feature. The higher the value, the greater the effect is. The results indicated that F0, duration and HNR of local and global features are very important,

while intensity has little effect. The important features (top 3 features) are summarized as follows:

- local features

1) *wh-HNR* of wh-word indicates the stress of Wh-word contribute a lot.

2) *Bt-F0_{slope}* and *Bt-dur* mean that the shape and duration of boundary tone are important.

3) *D-HNR* and *D-F0_{mean}* indicate that the stress different between wh-words and boundary tones are crucial.

- global features

S-LR and *S-FD_{mean}* illustrate that globe trends of the intonation are one of the major indexes in the discrimination.

4. Conclusion and Discussion

This paper has examined the prosodic differences between wh-questions and wh-declaratives, which have the same syntactic structure but different speech acts from the perspectives of both local and global prosodic features.

Based on the Chinese intonation components, we find that sentence final syllables attract the sentence stress in most cases in both speech acts. While in wh-questions, the chance that the stress falls on the wh-words is only 17.19%, however the stress will never locate on the wh-word in wh-statements. Therefore, if the intonation stress is on the wh-words, the sentence can be recognized as wh-question definitely.

Fisher discriminate analysis shows that local features made greater contribution than global features. When combining both global and local features, the best results could be obtained more than 85%. We summarized the top 3 features in this paper, and found HNR is a useful feature which relates to stress.

Unlike most of the modeling studies which only considered the global F0 features in the intonation discrimination, this paper introduced some local features which took into account the wh-words, the difference between wh-words and the following noun phrases and boundary tones.

In real applications, the performance of speech act discrimination is also dependent on that of speech recognition system. We should note that the data in the present study were small and well-controlled without any "noise" as found in real scenarios. There is no doubt that the performance must be reduced in spontaneous speech because of the "Noise" coming from aspects of environment, channels and so on. In fact, the encoding or decoding of speech acts is also multimodality and context dependent. Interlocutors use comprehensive features to get the intention of the speakers.

Nevertheless, this pilot study indicates that we should adopt local features based on the intonation grammar rather than the global features alone in the speech act discrimination. Prosodic structure relating to intonation, syntactic features and context features should be taken together into consideration in the future modeling with large data.

5. Acknowledgements

This research is supported by National Program on Key Basic Research Project (973 Program) under Grant 2013 CB329301, as well as the Innovation Program of Chinese Academy of Social Sciences "Key Laboratory of Phonetics and Speech Science".

6. References

- [1] X. Liu, A. Li, Y. J., "Functional Distributions of Interrogative Pronouns," in *IACL 2014-the 22nd International Association of Chinese Linguistics, May 2-4, America, Proceedings*, 2014, pp.180.
- [2] X. Liu, A. Li, Y. J., "Function and Stress of Sentences Bearing Interrogative Pronouns in Standard Chinese Dialogues," in *NCMMSC 2015-the 13th National Conference on Man-Machine Speech Communication, October 25-27, Tianjin, China, Proceedings*, 2015.
- [3] Y. Chao, "A Preliminary Study of English Intonation (with American variations) and its Chinese equivalents", *BIHP*, (The Ts'ai Yuan Pei Anniversary Volume), pp. 105-156, 1932.
- [4] Y. Chao, *A Grammar of Spoken Chinese*. Beijing: The Commercial Press, 1979.
- [5] M. Lin, Z. Li, "Focus and Boundary in Chinese Intonation," in *ICPhs 2011- the 17th International Congress of Phonetic Sciences, August 17-21, Hong Kong, China, Proceedings*, 2011, pp. 17-21.
- [6] D. R. Ladd, *Intonational Phonology*. Cambridge University Press, 2008.
- [7] F. Liu, Y. Xu. "Parallel Encoding of Focus and Interrogative Meaning in Mandarin Intonation," *PHONETICA*, vol.62, no. 2-4, pp.70-87, 2005.
- [8] K. Lambrecht, L. A. Michaelis, "Sentence Accent in Information Questions: Default and Projection," *Linguistics and philosophy*, vol.21, no.5, pp. 477-544, 1998.
- [9] N. Hedberg, J. M. Sosa, "The Prosody of Questions in Natural Discourse," in *SPEECH PROSODY 2002- the 1st Speech Prosody, April 11-13, France, Proceedings*, 2002.
- [10] J. Haan, "Speaking of Questions," *Netherlands Graduate School of Linguistics*, 2002.
- [11] A. Chen, "Interface between Information Structure and Intonation in Dutch WH-questions," in *SPEECHPROSODY 2006-the 3rd Speech Prosody, May 2-5, Dresden, Germany, Proceedings*, 2006.
- [12] F. Hu, "A Prosodic Analysis of Wh-words in Standard Chinese," *Studies of the Chinese Language*, vol.3, 2005.
- [13] J. Shen, "The Structures and Types of Mandarin Intonation," *Dialect*, vol.3, 1994.
- [14] M. Lin, *Experimental Studies of Mandarin Intonation*. Beijing: China Social Sciences Press, 2012.
- [15] M. Lin, "The Intonation of Interrogatives and Statements and Boundary Tone," *Studies of the Chinese Language*, vol.3, 2006, pp.364-376.
- [16] J. Yuan, "Mechanisms of Question Intonation in Mandarin," *Chinese spoken language processing, Springer Berlin Heidelberg*, pp.19-30, 2006.
- [17] J. Yuan, C. Shih, G. P. Kochanski, "Comparison of Declarative and Interrogative Intonation in Chinese," In *SPEECH PROSODY 2002-the 1st Speech Prosody, April 11-13, France, Proceedings*, 2002.
- [18] Z. Shiamizadeh, J. Caspers, N. O.Schiller, "Acoustic Correlates of Persian in-situ-wh-questions," in *ICPhS 2015-the 18th International Congress of Phonetic Sciences, August 10-14, Scotland, Proceedings*, 2015.
- [19] W. Lai, Y. Li, X. X, etc., "The Analysis and Modeling of the Intonation of Chinese Unmarked Questions," in *NCMMSC 2015-the 13th National Conference on Man-Machine Speech Communication, October 25-27, Tianjin, China, Proceedings*, 2015.
- [20] R. Sophie, "On the Acoustics of Wh-exclamatives and Wh-Interrogatives, Effects of Information Structure and Sex of Speaker," in *ICPhS 2015-the 18th International Congress of Phonetic Sciences, August 10-14, Scotland, Proceedings*, 2015.
- [21] O. J. Lee, "The Prosody of Questions in Beijing Mandarin," Ph.D. Dissertation. *The Ohio State University, Ohio*, 2005.
- [22] X. N. Shen, *The Prosody of Mandarin Chinese*. California: University of California Press, 1990.
- [23] F. Liu, D. Surendran and Y. Xu, "Classification of Statement and Question Intonations in Mandarin," in *SPEECHPROSODY 2006-the 3rd Speech Prosody, May 2-5, Dresden, Germany, Proceedings*, 2006, pp.5-25.
- [24] F. Liu, "Intonation Systems of Mandarin and English: a Functional Approach" Ph.D. Dissertation. *University of Chicago*, 2009.