

A Prosodic Analysis of Wh-words in Standard Chinese

Fang Hu

Department of Chinese, Translation and Linguistics
City University of Hong Kong
Fong_hwu@yahoo.com

Abstract

This paper is a phonetic study of prosody of wh-words in wh-questions, yes/no-questions and echo questions in Standard Chinese. The production data from 4 speakers show that there is a prosodic difference of wh-words and the corresponding VPs between different question types. Wh-words in wh-questions are the focus of sentence, whereas in yes/no-questions, VPs are the focus. The focused constituent is pitch accented so that its lexical tonal melody is retained and sometimes reinforced, while the lexical tonal melody of the corresponding unfocused constituent is compressed and sometimes reduced to a level tone. Speakers usually produce echo questions with a raised F0 register and an expanded F0 range of the wh-words to convey a meaning of surprise. However, no consistent data suggest that either duration or amplitude correlates with the focus vs. non-focus distinction.

1. Introduction

In Standard Chinese (hereafter, SC), wh-words may have two lexical functions: either as a wh-pronoun as in a content question, or as an indefinite pronoun as in a yes/no question or a conditional clause. The content question is marked with a particular particle 'ne', and the yes/no-question is marked with 'ma'. However, both the content-question marker 'ne' and the yes/no-question marker 'ma' are optional in SC, so a question sentence may be ambiguous due to the question marker drop. For example, 'shui' has two possible meanings, 'who' and 'anyone', and in (1a) and (1b) 'shui' only has one reading, 'who' and 'anyone' respectively, because the sentence is marked with different question particles, whereas in (1c), the sentence has two possible readings due to the lexical ambiguity of the wh-word 'shui'.

- (1) a. shui lai-le ne?
who come-ASP content-question marker
'Who is coming?'
b. shui lai-le ma?
anyone come-ASP yes/no-question marker
'Is there anyone coming?'
c. shui lai-le?
who/anyone come-ASP
'who is coming?/Is there anyone coming?'

In addition to this kind of lexical ambiguity, the wh-question in SC is also ambiguous with the echo question, because the wh-word in a wh-question is not moved to the beginning of a sentence as it should be in English, but stays in situ. Thus, a wh-question is not distinguishable from the corresponding echo question in which the echoed word is replaced with a wh-word in the same position.

In other languages, like Korean, this kind of ambiguity can be differentiated by prosodic features such as pitch accent, boundary tones, phonological phrasing, and so on (see [1] and references cited there). As is generally assumed, a questioned

constituent is focused by default, whereas an indefinite pronoun will never attract focus. One would expect that the wh-word is more prominent when used as a wh-pronoun in a wh-question than as an indefinite pronoun in a yes/no-question. In many languages, a high pitch accent is commonly used to convey focus (e.g., [2], [3], [4], [5]). One would also expect that intonational differences should be detected between the sentences like (1a) vs. (1b) and the two (1c) respectively in a tone language, SC, where pitch is mainly used to convey lexical meanings.

This paper is a phonetic analysis of prosody of wh-words in SC. It examines the prosodic features such as F0, duration, and amplitude of wh-words and other relevant constituents in wh-questions, yes/no-questions, and echo questions, to discover the prosodic differences of wh-words in these different sentence types from the production data.

2. Method

Material. Two wh-words, *shui* 'who/anyone' and *shen-me* 'what/anything', are chosen as the target wh-words, because they have the same lexical tonal melody MH, although *shui* is a monosyllable and *shen-me* is a disyllabic word. Sentences are designed so that the target wh-word occurs either at the beginning, in the middle, or at the end of a sentence. See (2):

- (2) a. shui lai-le
Who/anyone come-ASP
b. ni kan-jian shui lai-le
You watch-see who/anyone come-ASP
c. zhang-san mai-le shen-me
Name buy-ASP what/anything

The VPs followed the wh-word in (2a) and (2b) or preceded the wh-word in (2c) are also the target constituent under measure. The VP 'lai-le' has the same tonal melody MH with the adjacent wh-word, and the VP 'mai-le' has a similar tonal melody LH. The above sentences are in three-way ambiguous, namely a wh-question, a yes/no-question, or an echo question. And we distinguish two types of echo question: one is used when the listener is not sure of what the speaker said, the other is used when the listener knows what has been uttered but is surprised to hear it. In addition, as mentioned earlier, the sentence can be disambiguated by adding a proper question marker, namely a wh-question marker 'ne' or yes/no-question marker 'ma', which both have a high level tone. In this way, six different type sentences are designed for each sentence in (2).

Each type of question was putted into a corresponding dialogue: wh-questions or yes/no-questions were triggered by a proper answer, and echo questions were designed as a respond to a statement. In order to distinguish the two types of echo-question, additional statement was written in parenthesis next to the target question to provide the subjects with enough information.

Subjects and procedures. Four SC speakers in their late twenties were recorded: two male (M1 and M2) and two female (F1 and F2). M2, F1 and F2 were all born and raised up in Beijing, and M1 was born in Shandong, a Mandarin speaking area as well, and moved to Beijing with his family when he was 15. Now they are all graduate students in City University of Hong Kong.

Each designed dialogue was printed on a half-A4 size paper. A total of eighteen dialogues were pseudo-randomized so that one similar dialogue did not appear right after the other in sequence. For each dialogue, one speaker read the first sentence (a question or statement), and the second speaker read the second sentence (an answer corresponding to the question or an echo question corresponding to the statement). The eighteen dialogues were repeated five times by each subject. The recording was made in a soundproof booth with a Sony PCM-R700 Digital Audio Recorder and a Shure SM-58 Microphone. The speech was analyzed using Kay's CSL4400 speech analysis software. In each target utterance, F0 values of the lowest and highest points in the wh-word, the lowest and highest points in the target VP, and the question particle (if there is any) were measured; durations of the wh-word, VP, and question particle (if any) were measured; peak amplitude values of the wh-word, VP, and question particle (if any) were measured.

3. Results and Discussion

3.1. F0

Results show that there is an intonational difference of wh-words and the corresponding VPs between the wh-question (WHQ) and yes/no-question (YNQ). In a WHQ, the wh-pronoun is focused and consequently accented, so that the lexical tonal melody MH is retained and sometimes reinforced while the lexical tonal melodies of the corresponding VP MH and LH are somewhat weakened and sometimes leveled to a M (or H) and L tone respectively. Typically, in a WHQ, the highest F0 point on wh-words is much higher than that on the corresponding VPs, and the pitch range of wh-words is considerably greater than that of VPs, irrespective of the position of wh-words in a sentence; in a YNQ, the VP has a much higher highest F0 point and a much greater pitch range.

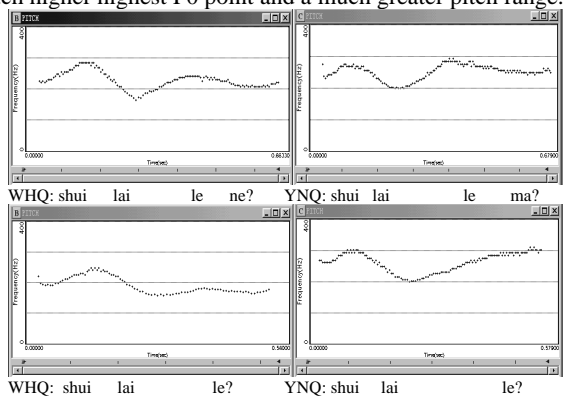


Figure 1: F0 contours of sentences 'shui lai-le ne?' (top left), 'shui lai-le ma?' (top right), 'shui lai-le?' (WHQ) (bottom left), and 'shui lai-le?' (YNQ) (bottom right) by the speaker F2.

Figure 1 is an example of F0 contours of F2 when producing WHQs and YNQs with or without particles respectively (wh-words in sentence-initial position).

Results of the detailed lowest and highest F0 values of wh-words, the corresponding VPs, and the question particles (if any), with their Standard Deviations in parentheses right after the mean, are summarized through table 1 to table 6. From the tables we can see that all speakers tend to retain or reinforce the lexical tonal melody, MH of wh-words in WHQs while tend to reduce it in YNQs. And this is also true for the VPs. Results suggest that in SC, wh-words are focused while VPs are unfocused in WHQs, whereas wh-words are unfocused while VPs are focused in YNQs. The focused constituent has a pitch accent so that its lexical tone is retained and sometimes reinforced, whereas the lexical tonal melody of the unfocused constituent is compressed or even reduced to a level tone. However, the F0 range expansion and compression caused by focus and non-focus effect respectively may have inter-speaker and/or intra-speaker differences. And the F0 range compression of unfocused constituents is, in general, considerably greater after focus than before focus (cf.: [6]).

Comparing with the highest F0 point of the preceding VP, particles have a similar or a little bit lower F0 value when wh-words are not in a sentence-final position. However, it is interesting to find that particles always have a higher F0 value than the highest point of the preceding wh-words when wh-words are in the sentence-final position.

One more thing that should be pointed out is that creaky voice appears, more often on the VP and sometimes on the wh-word, when wh-words are in a sentence-final position, because the citation tone of the verb 'mai' is low dipping, which is realized as low in connected speech, and this low tone is often realized as the creaky voice in SC. If the constituent is unfocused, the whole VP and sometimes the whole wh-word may be creaky, whereas if the constituent is focused, it may begin with a creaky voice, but always keeps its rising tonal melody and ends with a normal voice at a certain F0 range. In table 5 and 6, the gray-shadowed cells indicate that one or two measured tokens are creaky and the given F0 values are the mean of the rest tokens, and if there are more than two creaky tokens, the cell of mean F0 value is marked as 'creaky'.

Table 1: Mean F0 values (in Hz, n=5), with SDs in parentheses right after the means, of wh-words and the following VPs in WHQs and YNQs with particles (wh-words in a sentence-initial position)

Sent.Type & subjects		Wh-word		VP		Particle
		L	H	L	H	
WHQ	M1	146(4)	178(4)	151(8)	174(12)	172(12)
	M2	135(7)	150(11)	127(9)	136(11)	132(10)
	F1	216(16)	289(21)	227(16)	229(19)	220(19)
	F2	216(11)	305(67)	200(22)	266(41)	238(30)
YNQ	M1	137(6)	151(8)	136(14)	179(12)	171(14)
	M2	140(8)	154(8)	127(4)	149(6)	132(3)
	F1	246(25)	285(13)	223(19)	266(13)	239(34)
	F2	256(15)	288(16)	211(12)	295(14)	260(13)

Table 2: Mean F0 values (in Hz, n=5), with SDs in parentheses right after the means, of wh-words and the following VPs in the ambiguous WHQs and YNQs (wh-words in a sentence-initial position)

Sent. Type & subjects		Wh-word		VP	
		L	H	L	H
WHQ	M1	146(10)	182(6)	156(13)	163(16)
	M2	131(7)	156(10)	128(10)	131(9)
	F1	205(20)	268(29)	180(14)	190(18)
	F2	210(120)	285(29)	178(16)	193(14)
YNQ	M1	138(7)	151(6)	136(11)	182(16)
	M2	132(2)	146(2)	124(5)	149(4)
	F1	240(19)	278(26)	228(22)	289(29)
	F2	237(16)	268(17)	212(8)	307(9)

Table 3: Mean F0 values (in Hz, n=5), with SDs in parentheses right after the means, of wh-words and the following VPs in WHQs and YNQs with particles (wh-words in a sentence-middle position)

Sent. Type & subjects		Wh-word		VP		Particle
		L	H	L	H	
WHQ	M1	152(9)	184(9)	147(3)	168(5)	160(10)
	M2	133(7)	146(9)	125(7)	131(10)	127(8)
	F1	210(27)	261(16)	196(13)	204(22)	207(13)
	F2	201(16)	271(21)	169(27)	225(35)	203(37)
YNQ	M1	143(8)	158(8)	144(10)	182(11)	177(8)
	M2	131(3)	147(6)	123(7)	147(8)	132(9)
	F1	232(3)	275(7)	216(10)	253(14)	236(23)
	F2	229(25)	264(35)	186(21)	273(35)	242(35)

Table 4: Mean F0 values (in Hz, n=5), with SDs in parentheses right after the means, of wh-words and the following VPs in the ambiguous WHQs and YNQs (wh-words in a sentence-middle position)

Sent. Type & subjects		Wh-word		VP	
		L	H	L	H
WHQ	M1	138(6)	171(3)	135(10)	135(10)
	M2	126(5)	137(10)	121(6)	129(4)
	F1	190(8)	261(13)	170(5)	176(9)
	F2	198(18)	262(21)	187(30)	209(33)
	F2	198(18)	262(21)	187(30)	209(33)
YNQ	M1	138(11)	161(7)	142(7)	188(6)
	M2	130(5)	146(5)	124(2)	150(5)
	F1	240(22)	285(25)	221(35)	302(39)
	F2	255(10)	272(10)	222(11)	294(6)

Table 5: Mean F0 values (in Hz, n=5), with SDs in parentheses right after the means, of wh-words and the preceding VPs in WHQs and YNQs with particles (wh-words in a sentence-final position)

Sent. Type & subjects		VP		Wh-word		Particle
		L	H	L	H	
WHQ	M1	115(11)	134(12)	113(6)	152(6)	177(14)
	M2	114(5)	122(5)	116(3)	133(6)	136(6)
	F1	Creaky	202(10)	178(19)	224(17)	253(27)
	F2	133(17)	180(9)	138(13)	231(27)	277(34)
YNQ	M1	90(4)	136(11)	106(2)	106(2)	118(2)
	M2	Creaky	128(4)	118(2)	119(3)	117(3)
	F1	Creaky	215(24)	192(34)	213(23)	238(15)
	F2	138(14)	180(11)	152(7)	192(25)	212(23)

Table 6: Mean F0 values (in Hz, n=5), with SDs in parentheses right after the means, of wh-words and the preceding VPs in the ambiguous WHQs and YNQs (wh-words in a sentence-final position)

Sent. Type & subjects		VP		Wh-word	
		L	H	L	H
WHQ	M1	107(7)	136(22)	107(4)	157(22)
	M2	115(5)	124(6)	115(8)	142(7)
	F1	Creaky	193(9)	165(12)	206(9)
	F2	144(12)	184(13)	155(10)	264(26)
YNQ	M1	Creaky	134(8)	98(1)	102(8)
	M2	Creaky	128(7)	117(3)	124(3)
	F1	Creaky	232(12)	172(12)	172(13)
	F2	Creaky	237(7)	Creaky	Creaky

I have shown above that the tonal behavior of wh-words in WHQs is different from that in YNQs. And this suggests the prosodic disambiguation of the ambiguous sentences like (1c) between a WHQ and YNQ reading in terms of intonation. Now we come to the discussion of tonal behaviors of wh-words in echo-questions. In this study, a distinguishing is made between two types of echo questions: one is used when the listener is not sure of what the speaker said, the other is used when the speaker is surprised to hear what the speaker said. Mean F0 values, with SDs in parentheses right after the means, of wh-words and the corresponding VPs in echo questions containing the emotion of surprise are summarized below in table 7.

Table 7: Mean F0 values (in Hz, n=5), with SDs in parentheses right after the means, of wh-words and the corresponding VPs in echo questions

Wh-words sent.-initial		Wh-word		VP	
		L	H	L	H
Echo	M1	153(21)	206(26)	193(24)	236(41)
	M2	146(5)	163(8)	145(9)	153(6)
	F1	211(28)	300(33)	232(29)	245(32)
	F2	217(28)	397(54)	308(36)	406(54)
Wh-words sent.-mid.		Wh-word		VP	
		L	H	L	H
Echo	M1	173(8)	217(15)	197(13)	234(25)
	M2	142(5)	168(5)	149(4)	152(5)
	F1	213(28)	327(30)	242(33)	260(38)
	F2	254(33)	392(70)	307(48)	396(110)
Wh-words sent.-final		VP		Wh-word	
		L	H	L	H
Echo	M1	132(11)	142(12)	138(15)	257(23)
	M2	121(2)	124(2)	127(3)	190(11)
	F1	142(27)	194(6)	185(16)	356(24)
	F2	193(17)	206(11)	186(9)	434(57)

Compared with the data of normal WHQs in table 2, 4 and 6, wh-words in echo questions with the emotion of surprise have a much greater pitch range, except of one case where the F0 range is smaller when the speaker M2 produce the wh-word in the sentence-initial position. In addition to the pitch range difference, wh-words in echo questions with the emotion of surprise tend to have a higher F0 register, namely both the highest and the lowest F0 point are usually (although not always) much higher than those in normal WHQs, and this is also true for the exceptional case in terms of F0 range. Moreover, the unfocused VPs also tend to have a much higher

F0 register, especially when the wh-word is not in a sentence-final position, and in the speakers of M1 and F2, the highest F0 point of VPs is even much higher than that of the preceding wh-words. As for the case where wh-words are in the sentence-final position, although the highest F0 point of the unfocused VP is sometimes even lower than that in normal WHQs, the creaky voice at the lowest F0 point of VPs, which is common in normal WHQs, only occur twice in the speaker F1, which also indicates the raise of the F0 register. All these suggest that in an echo question, speakers generally tend to raise the pitch register of the whole sentence and expand the F0 range of wh-words to express the emotion of surprise. And the data also suggest that although wh-words in WHQs are the focus and consequently pitch accented, wh-words in echo questions need extra pitch accent to convey the meaning of surprise.

However, for the echo questions used when the listener is not sure of what the speaker said, no consistent F0 data is found in this study. Roughly speaking, speakers sometimes use an F0 pattern more similar to that of a normal WHQ, but sometimes use an F0 pattern more similar to that of an echo question conveying the emotion of surprise.

3.2. Duration and amplitude

In SC, wh-words in WHQ are the focus of a sentence. Focused constituents are perceptually more prominent than the unfocused ones, which may correlate with pitch accent (F0) (as shown above by the data), duration, and/or amplitude. In this section, I will discuss the duration data and amplitude data obtained in this study. For the duration, to limit the size of comparisons, only the duration data of wh-words in three types of ambiguous sentences, namely WHQs and YNQs without question particles and the echo questions containing a meaning of surprise, are presented here. And durations of wh-words in WHQs are compared directly with the corresponding data in the YNQs and echo questions. The mean durations, with SDs in parentheses right after the means, of wh-words in WHQs and the corresponding YNQs and echo questions are summarized in table 8.

Table 8: Mean durations (in millisecond, $n=5$), with SDs in parentheses right after the means, of wh-words in WHQs, YNQs, and echo questions

Wh-words in sent.-initial position			
Subj.	WHQ	YNQ	Echo
M1	198(11)	185(16)	204(14)
M2	296(9)	254(16)	221(24)
F1	187(22)	159(11)	234(27)
F2	209(7)	189(12)	321(39)
Wh-words in sent.-middle position			
Subj.	WHQ	YNQ	Echo
M1	148(13)	142(27)	144(16)
M2	185(21)	184(20)	173(12)
F1	184(16)	157(7)	212(13)
F2	195(19)	133(6)	259(38)
Wh-words in sent.-final position			
Subj.	WHQ	YNQ	Echo
M1	298(8)	256(6)	255(11)
M2	299(8)	278(17)	241(34)
F1	422(28)	351(24)	280(32)
F2	367(43)	308(35)	369(23)

From the table we can see that the duration data lack consistency. It is true that wh-words in WHQs have longer mean durations than wh-words in YNQs, but in most cases, the duration difference is not significant if referring to the standard deviations. Only three cases marked with gray shadows in the table suggest a considerably difference (Note that the difference, 442 vs. 351ms, of speaker F1 when wh-words are in the sentence-final position is not taken into the account, because F1 tends to deliberately lengthen the schwa sound in the final syllable of wh-words when producing this type of WHQs). As for the difference between normal WHQs and echo questions, data are also not consistent. Speakers sometimes lengthen, but sometimes shorten the duration of wh-words in echo questions.

The peak amplitudes of wh-words, VPs and particles (if any) are measured in this study, but unfortunately, no consistent data is detected. Speakers sometimes produce the focused constituent with higher peak amplitude, sometimes produce the unfocused constituent with higher peak amplitude, and sometimes even produce the question particle with the highest amplitude.

4. Conclusion

The results of this paper can be summed up as follows:

(1) Wh-words in WHQs are the focus of sentence, whereas in YNQs, wh-words are not the focus, instead the corresponding VPs are the focus, irrespective of the position of wh-words in the sentence.

(2) Data from the all four speakers show that focus is manifested with pitch accent, an effect of intonation. The focused constituent, either wh-words or VPs, retains and sometimes expands its lexical tonal melody, while the corresponding unfocused constituent compresses and sometimes reduces its lexical tonal melody to a level tone.

(3) The presented production data suggest that the possible readings of ambiguous question sentences as WHQs, YNQs and one type of echo questions can be disambiguated by the prosodic differences summarized in (1) and (2).

(4) No consistent data of duration and amplitude are detected which suggest a correlation with the focus vs. non-focus distinction, although speakers sometimes do produce the focused wh-words in WHQs with considerably longer duration than produce the unfocused wh-words in YNQs.

5. References

- [1] Jun, S.; Oh, M., 1996. A prosodic Analysis of three types of wh-phrase in Korean. *Language and Speech* 39(1), 37-61.
- [2] Pierrehumbert, J., 1980. *The phonology and phonetics of English intonation*. Ph.D. Dissertation, MIT.
- [3] Pierrehumbert, J.; Beckman, M., 1988. *Japanese tone structure*. Cambridge, MA: MIT Press.
- [4] Selkirk, E., 1995. Sentence prosody: Intonation, stress, and phrasing. In *Handbook of phonological theory*, J. Goldsmith (ed.). Oxford: Blackwell, 550-569.
- [5] Ladd, D., 1996. *Intonational phonology*. Cambridge, U.K.: Cambridge University Press.
- [6] Gårding, E., 1987. Speech act and tonal pattern in Standard Chinese: Constancy and Variation. *Phonetica* 44: 13-29.