Sentence Prosody and Wh-indeterminates in Taiwan Mandarin

Yu-Yin Hsu, Anqi Xu

1Department of Chinese and Bilingual Studies, Hong Kong Polytechnic University, Hong Kong
2Department of Speech, Hearing and Phonetic Sciences, University College London, UK

yu-yin.hsu@polyu.edu.hk, a.xu.17@ucl.ac.uk

Abstract

We report results of a speech production experiment about the intonation of three sentence types in Taiwan Mandarin, and discuss our results with implications for focus acoustics, and semantic-syntactic theories of sentence final particles and wh-indeterminates. Wh-indeterminates refer to wh-phrases that are ambiguous between interrogative and indefinite readings. In Mandarin, different interpretations of wh-indeterminates are not morphologically marked, but can be disambiguated in specific sentence contexts marked by sentence final particles. In this study, we systematically examined the intonation of wh-questions and yes/no questions by using declarative sentences as the baseline. The results show that both wh- and yes/no questions exhibit F0 prominence, and lengthening effects on regions containing sentence-final particles and wh-phrases, but the effects were stronger in wh-questions. Examining the duration and F0 range, we found that wh-phrases and sentence final particles together formed specific acoustic patterns to distinguish questions from declarative sentences. The findings suggest that the prosodic organization interacts with other internal structural organization.

Index Terms: sentence prosody, wh-indeterminate, sentence final particle, focus prosody, Taiwan Mandarin

1. Introduction

Cross-linguistically, one interesting and important research question has been on how linguistic ambiguity is resolved and how tone languages express intonation of different types of sentences. In this study, we examined the acoustic nature of a special lexical category – wh-indeterminates (i.e., wh-phrases that are ambiguous between interrogative and indefinite readings but are disambiguated in specific sentence types), which has been attested in many languages ([1], [2]). Mandarin wh-phrases like shénme ‘what’ can be interpreted as an interrogative in a wh-question (1a), or an indefinite in a yes/no question (1b). Previous syntax-semantics studies have identified the occurrence of sentence final particles (henceforth SFPs) ma for yes/no questions as one of such wh-indeterminate-licensing contexts ([3], [4]). While most linguistic studies have focused on sentences like (1a-b), we notice that in declarative sentences like (1c) with the SFP ba (indicating weak epistemic judgment [5]), shénme therein is also interpreted as an indefinite noun.

(1) a. Zhāngsān mài-le shénme ne? Zhāngsān mài-le shénme ma?
   ‘What did Zhangsan buy?’ ‘Did Zhangsan buy something/anything?’
   Zhāngsān mài-le shénme ba.
   Zhāngsān mài-le shénme ba.
   Zhāngsān mài-le shénme ba.
   Zhāngsān mài-le shénme ba.

Examples in (1) show that Chinese wh-indeterminates not only are lexically ambiguous but are also relevant to structural ambiguity, given that SFPs often are not obligatory in Chinese.

Considering prosody as one of the disambiguation devices [6], some studies have reported that wh-indeterminates while functioning as wh-interrogatives manifest more acoustic prominence than wh-indeterminates in languages like Korean [7], Japanese [8], [9] [10], and German [11]. Yet, some reported that no acoustic differences on wh-indeterminates that distinguished interrogative from indefinite readings (e.g., [12]).

Different results were also reported for wh-indeterminates in Mandarin. Hu [13] studied wh-subject shei ‘who’ and shénme ‘what’ and reported that Mandarin speakers expressed wh-interrogatives acoustically different from wh-indeterminates, i.e., wh-phrases had higher mean F0 in wh-questions, and the verb phrase of a sentence showed higher mean F0 in yes/no questions. In this study, only descriptive statistics were reported for mean F0, duration and amplitude (with SD), and some inter-participant differences were found. For Taiwan Mandarin (henceforth TwM), Shyu and Tung [14] reported two different findings; first, based on eight tokens (from a corpus [15]), they reported some differences between wh-interrogatives and indefinites, but the syntactic and phonetic contexts where these eight tokens occurred were different; second, their production study showed that speakers did not acoustically disambiguate wh-indeterminates; however, since participants responded to the same two items for one context, it is difficult to draw a general conclusion for TwM. Thus far, the findings about Mandarin wh-indeterminates seem not yet to be conclusive.

Most of the theoretical studies about Chinese sentence prosody assumed the sentence final position being the locus of nuclear stress (cf. [16]). Therefore, some researchers proposed that sentence final position does not attract acoustic prominence, since Chinese languages use lexical tones and clausal types are already expressed by syntax ([17], [18]), while some studies reported that wh-questions had higher overall F0 contours than that of statements ([19], [20]).

With respect to focus prosody, much fewer studies on the wh-phrases themselves that induce focus interpretations [21]; most studies about Chinese focus prosody have examined the acoustic prominence on the answer to a wh-question (with no SFP) (e.g., [22] [23]). While different Chinese varieties and languages may use different acoustic devices to express focus, it has been consistently reported for Chinese languages that focus units show on-focus F0 rising and lengthening [24] [25] [26] [27]. And, to the best of our knowledge, even less research has investigated the prosodic function of SFPs in Chinese languages [28], and no research is about the impacts of SFPs on focus prosody and sentence intonation. Given that the majority of prosody work has been on Beijing Mandarin, and that the phenomenon of wh-indeterminates (e.g., (1)) provides us a new...
and ideal acoustic context to systematically examine theoretical claims and experimental findings, we used this context to first, study how the sentential prosodic organization interacts with the system of focus marking and the influences from syntax [29]; second, to contribute to the research field with an interesting but relatively understudied Mandarin variety, TwM.

We used a speech production experiment to study whether and how TwM users use prosody to disambiguate *wh*-indeterminates, to examine whether sentence final position in specific syntactic constructions can bear focus acoustic prominence, and to explore whether and how SFPs signal the sentence types and interact with the system of focus marking.

## 2. Method

### 2.1. Stimuli

We examined the prosody of *wh*-questions and yes/no questions containing *wh*-indeterminates by using declaratives as the baseline, as shown in Table 1. Target sentences were constructed in the same tonal format, i.e., a monosyllabic subject pronoun in Tone3 (i.e., 彼 ni3 ‘you’ and 我 wo4 ‘I’) followed by a disyllabic auxiliary 可以 ke4yi3 ‘can’, and they were followed by a disyllabic helping verb or a disyllabic sentential adverb in the same Tone1-Tone2 sequence (e.g., 幫忙 bang‘mang2 ‘help’, or 今年 jin4nian2 ‘this year’). The main verb immediately before the target *wh*-phrase was a Tone4 monosyllabic verb (e.g., 帶 dai4 ‘bring’, 要 ya4 ‘ask for’). These choices were due to lexical limitation while we considered the need of having enough parallel stimuli among sentence types and of maintaining the same tonal contexts that hosted the *wh*-indeterminates. Right after the main verb was the *wh*-phrase target and it was immediately followed by a Tone1 SFP indicating one of the three sentence types (i.e., 呢 ne for *wh*-Q, 嗎 ma for yes/no-Q, and 吧 ba for declaratives).

**Table 1: Examples of target sentences with ‘what’**.

<table>
<thead>
<tr>
<th>Sentence Types</th>
<th>Format of Target Sentences</th>
<th>T3</th>
<th>T2-T3</th>
<th>T1-T2</th>
<th>T4</th>
<th>Wh</th>
<th>T1</th>
<th>SFP</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Wh-Q</em></td>
<td>(Ad)‘can’</td>
<td>你</td>
<td>可以</td>
<td>幫忙</td>
<td>帶</td>
<td>什麼</td>
<td>呢?</td>
<td></td>
</tr>
<tr>
<td><em>Yes/no Q</em></td>
<td>(Ad)‘can’</td>
<td>你</td>
<td>可以</td>
<td>幫忙</td>
<td>帶</td>
<td>什麼</td>
<td>呢?</td>
<td></td>
</tr>
<tr>
<td><em>Declarative</em></td>
<td>(Ad)‘can’</td>
<td>我</td>
<td>可以</td>
<td>幫忙</td>
<td>帶</td>
<td>什麼</td>
<td>吧.</td>
<td></td>
</tr>
</tbody>
</table>

Four *wh*-phrases (she1 ‘who’, shen.me ‘what’, na.li ‘where’, and shen.me-dong.xi ‘what thing’) were used to construct each type of sentences with five versions of Tone4 verbs. In total, 60 target sentences (3 sentence types x 4 *wh*-phrases x 5 verb) and 40 filler sentences (in different sentence structure with no SFPs) were used. Each target trial consisted of a pre-recorded context (25 character long) (A), and a target sentence (B) that participants used to respond. The contexts were pre-recorded by a female speaker of Mandarin from Taiwan, as shown in (2-4).

(2) *Wh*-question

A: 我等會兒要出去買飯。有沒有人需要我順路辦事或者帶飯的？ *I am about to go to buy my meal. Does anyone need me to run simple errands or to buy a meal on the way?*  
B: 你可以幫忙帶什麼呢？ *Can you help me bring something?*

(3) *Yes/no question*  
A: (The leading context same as the one in the *wh*-question)  
B: 你可以幫忙帶什麼嗎？ *Can you help bring something?*

(4) Declarative sentence  
A: 星期六我們要去大湖邊野餐。有沒有人有空幫忙帶些東西呢？ *This Saturday we are going on a picnic by the big lake. Can anyone help to bring something for the picnic?*  
B: 我可以幫忙帶什麼呢。 *I probably can help to bring [us] something.*

### 2.2. Participants

The ethics approval and basic demographic information were obtained before the experiment. 10 female native Mandarin speakers born and raised in Taiwan (mean age ± SD: 20.1 ± 0.83 years), who were university students in Hong Kong, joined our study. None reported any history of hearing problems. Based on a 7-point self-report scale, 4 reported non-fluency in Taiwanese Min (1-3 points), and 6 reported intermediate fluency (4-5 points); As for Cantonese, 8 reported non-fluency (1-3 points), and 2 reported intermediate fluency (4-5 points). Participants were paid HK$50 after the experiment.

### 2.3. Procedure

The experiment was conducted in Hong Kong in a sound-attenuated lab with a Focusrite Scarlet 212 sound interface, and a Telefunken M-80 dynamic microphone calibrated to measure intensity. Each participant first signed a consent form and filled out a background questionnaire. Participants were seated in front of a computer screen and wore headphones. Stimuli were presented one at a time (self-paced) on the screen. The order of trials was pseudo-randomized, such that no similar target item occurred immediately adjacent. Participants were asked to first listen to the leading context, and then read the target sentence aloud as casually and naturally as possible; no instructions were given regarding emphasis. Participants produced each sentence twice, and additional repetitions were allowed in cases of mispronunciation or hesitation. Productions were recorded in .wav format at a sampling rate of 44.1 kHz with 16-bit quantization. There were three practice trials before the main trials. Each experiment lasted about 30 minutes.

### 2.4. Acoustic measurements

The acoustic measurements were generated by ProsodyPro 5.7.6 [30] for duration, fundamental frequency (F0) range and time normalized F0. Syllable boundaries were determined by using both visual (the waveform and spectrogram) and auditory information. The vocal pulses were manually checked and corrected when there were pitch halving or doubling and creaky voice. F0 was time-normalized across tokens by dividing each syllable into 10 intervals equal in time and the trimmed F0 values were calculated. The graphical analysis of F0 was performed by Smoothing Spline ANOVA (SS-ANOVA) [31]. When the Bayesian 95% confidence intervals indicated by transparent ribbons around the means do not overlap, the F0 curves are significantly different from each other.

Linear Mixed Effects models were conducted on duration and F0 range using the lme4 package [32] in R [33]. The model first included random intercepts for item and speaker and also by-speaker, by-ITEM, by-SPEAKER-ITEM-interaction random slopes for SENTENCE TYPE (declaratives, *wh*- and yes/no questions). SENTENCE TYPE was then added as a potential fixed effect. The significance of the main effect was evaluated by

3951
3. Results

3.1. F0 contours

The time-normalized F0 contours of statements, wh- and yes/no questions with four wh-phrases are in Figure 1. In general, wh-questions and yes/no questions were realized similarly in terms of F0, which contrasted with declarative sentences. The difference between statements and questions emerged quite early (at around the first three syllables), that is, the statements had higher F0 than the questions. Such tendency was more prominent in sentences with ‘who’, ‘where’ and ‘what thing’, as for sentences with ‘what’, the differences was prominent at the first syllable.

![F0 contours](image)

Figure 1: SS-AANOVA plots of F0 contours.

The most noticeable divergence across sentence types was found toward the end of the sentences, where the F0 contours differed greatly in the region of wh-phrases and SFPs (marked by the inner boxes inside of the SS-AANOVA plots). Specifically, both the wh- and yes/no questions had overall higher F0 than statements on the wh-phrases, and more complex F0 transitions in the syllable of SFPs. The more it was closer to the end of a sentence, the more prominent the modulation of F0 range was. While the two types of questions largely overlapped with each other across the four wh-phrase conditions, interestingly, in sentences with ‘who’ and ‘what thing’, the F0 of wh-questions was overall significantly higher than that of yes/no questions. Yet, in sentences with ‘where’, both questions showed lower F0 toward the end; the F0 curve of wh-questions was slightly lower than that of yes/no questions.

3.2. Duration and F0 range

Figure 2 shows the mean and the distribution of the duration and the F0 range of each syllable of wh-phrases and of SFPs in three sentence types. The asterisks indicate significant difference between conditions in post-hoc comparisons. One interesting finding is that no matter how many syllables the wh-indeterminates have, the duration and the F0 range differ significantly between statements and questions mainly in the last syllable of the wh-indeterminates and in SFPs.

![Duration and F0 range plots](image)

Figure 2: Duration and F0 range of wh-indeterminates and sentence final particles.

The results of mixed models are summarized in Table 2. In sentences with ‘who’, the main effect of SENTENCE TYPE was significant on the duration of the SFPs and of the last syllable of wh-phrases. The duration of SFPs and the syllable before SFPs in statements was significantly different from that of questions. The main effect of SENTENCE TYPE was also significant on the F0 range of the SFPs, and marginally significant on the last syllable of wh-phrases. Wh-questions had narrower F0 range than statements and yes/no questions in SFP.

In sentences with ‘what’, the analysis confirmed that the main effect of SENTENCE TYPE was significant on the duration of SFPs and on the last syllable of wh-phrases. The post-hoc comparison showed that the duration of the SFPs in wh- and yes/no questions were longer than statements, whereas the duration of the syllable before the SFPs in questions was shorter than that in statements. However, the two questions did not significantly differ in terms of duration. The main effect of SENTENCE TYPE was marginally significant on F0 range of SFPs and was significant on the last syllable of wh-phrases. The F0 range of wh-question and yes/no question was narrower than
statements in the syllable right before the SFPs, while the overall F0 range of the two types of questions was similar.

Regarding sentences with ‘where’, the main effect of SENTENCE TYPE was significant on the duration of the SFPs and of the last syllable of wh-phrases. The duration of questions was longer than that of statements in the SFPs. Noticeably, the duration of the syllable before the SFPs was the longest in statements, and such syllables in wh-questions were longer than those in yes/no questions. The main effect of SENTENCE TYPE was significant on the F0 range of the SFPs and of the last syllable of wh-phrases. Questions had narrower F0 range than the statements in the SFPs and in the syllable before SFPs.

In sentences with ‘what thing’, the main effect of SENTENCE TYPE was significant on the duration of the SFPs and on the last syllable of wh-phrases and on the second syllable of wh-phrase (‘me’ region). Yes/no questions had duration longer than wh-questions in the second syllable of wh-phrases. Compared with statements, lengthened SFPs and shortened final syllable of wh-phrases were found in both questions. The main effect of SENTENCE TYPE was significant on the F0 range of SFP and of the last syllable of wh-phrases, where the differences was found significantly between questions and statements, but not between the two question types.

Table 2: The main effect of SENTENCE TYPE.

<table>
<thead>
<tr>
<th></th>
<th>Duration</th>
<th>F0 range</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>y2</td>
<td>df</td>
</tr>
<tr>
<td>Who</td>
<td></td>
<td></td>
</tr>
<tr>
<td>shen</td>
<td>9.898</td>
<td>2</td>
</tr>
<tr>
<td>SFP</td>
<td>20.107</td>
<td>2</td>
</tr>
<tr>
<td>What</td>
<td></td>
<td></td>
</tr>
<tr>
<td>shen</td>
<td>3.267</td>
<td>195</td>
</tr>
<tr>
<td>me</td>
<td>16.061</td>
<td>2</td>
</tr>
<tr>
<td>SFP</td>
<td>20.600</td>
<td>2</td>
</tr>
<tr>
<td>Where</td>
<td></td>
<td></td>
</tr>
<tr>
<td>shen</td>
<td>3.028</td>
<td>220</td>
</tr>
<tr>
<td>li</td>
<td>22.698</td>
<td>2</td>
</tr>
<tr>
<td>SFP</td>
<td>27.992</td>
<td>2</td>
</tr>
<tr>
<td>What thing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>shen</td>
<td>.321</td>
<td>852</td>
</tr>
<tr>
<td>me</td>
<td>6.553</td>
<td>2</td>
</tr>
<tr>
<td>dong</td>
<td>.423</td>
<td>810</td>
</tr>
<tr>
<td>xi</td>
<td>20.866</td>
<td>2</td>
</tr>
<tr>
<td>SFP</td>
<td>22.291</td>
<td>2</td>
</tr>
</tbody>
</table>

4. Discussion and Conclusion

We studied whether Taiwan Mandarin speakers prosodically distinguish wh-interrogatives from indefinites, and whether sentential prosodic organization interacts with the systems of focus marking and syntax (as represented by the occurrence of SFPs). Our results show that the mechanism of sentential prosodic organization not only considers lexical tones, but also sentence types and focus marking, as target sentences across sentence types bear identical lexical tones, and yet, consistent patterns were found to distinguish states from questions, and specific patterns of wh-focus were also observed.

Concerning focus marking, we observed that wh-phrases in wh-questions showed significantly higher F0 than wh-phrases in declaratives across all four types of wh-phrases. This finding is not only in line with previous studies on focus prosody, but also indicates that TwM speakers acoustically distinguished wh-interrogative (in wh-questions) from wh-indefinite phrases (in declaratives). Also worth noting is that while the overall F0 contours of wh-questions and yes/no questions were similar, the F0 of wh-phrases were significantly higher than that of yes/no questions in ‘who’, ‘what’, and ‘what thing’, and only when the wh-phrase was ‘where’, the F0 of wh-phrases was lower in wh-questions than when it was in yes/no questions, although such a difference did not reach statistical significance. This reversed pattern might be due to the lexical tone of ‘where’ which ends in a low tone. One possibility would be that unlike high tones units that mark focus through raising the F0, low tone units may do so by lowering the F0. We leave this point for future study.

Our results also show that sentence final position can bear acoustic prominence. The overall F0 analyses across sentence types show that declaratives exhibit sentence initial prominence with a gradual lowering toward the end of the sentence, whereas both questions more dramatically raise the F0 on wh-phrases and on the SFPs. This sentence-final rising of F0 observed in both questions may seem to suggest that TwM speakers do not distinguish the two readings of wh-indeterminate. However, considering that they did prosodically distinguish them in wh-questions and statements, we think that this sentence final similarity of both types of questions may be due to different reasons. For wh-questions, it is expected, because the wh-phrases are the second unit to the last in each target sentence. For yes/no questions, although it is not expected to see a rise of F0 of wh-phrases expressing the indefinite reading in yes/no questions, the syllable immediately following the wh-indefinite is SFP indicating the yes/no question status. Thus, the sentence final F0 rising may be due to the sentential focus marked by the SFP of yes/no questions (cf. similar but less prominent F0 rising, while comparing it with the SFP of wh-questions).

In addition to F0, our results show very interesting and consistent patterns of duration and F0 range, across four wh-phrase types. Unlike the on-focus lengthening effect reported in previous studies, in our study the last syllable of all four types of wh-phrases in both questions was significantly shorter than that syllable in declaratives. This may seem rather surprising until we looked at the SFPs (immediately followed the target wh-phrases), i.e., SFPs in both questions were significantly longer than the SFPs in declaratives. Considering the duration patterns of wh-phrases and SFPs, we think that these two units may cooperate in marking focus prosody. Interestingly, the patterns of F0 range were also consistent across all four types of wh-phrases. The F0 range of statements was wider than both types of questions in the SFPs and in the last syllable of wh-phrases. These suggest that the occurrence of SFPs explicitly defines the sentence types, and while maintaining the lexical tones, the syntactic information indicated by SFPs requires the prosodic organization to comply with focus marking, as shown by the acoustic patterns of Ton1 SFPs across sentence types.

In sum, our results suggest that the internal information at different structural levels interact with the sentential prosody. We expect to see similar interactions in other Mandarin varieties, and would like to testify such effects in different groups of Mandarin speakers (e.g., male speakers, and learners of Mandarin), and other (Sinitic) languages in the future, to further advance our understanding of the language system.

5. Acknowledgements

This research was supported by the University Central Research Grant (G-YBR8), and by the departmental research fund (4-ZZQ) from Chinese and Bilingual Studies at PolyU. We thank Ka-Wai Chan, Tsz-Shan Lo, Xia Wang, and Ka-Keung Lee for technical support. Mistakes remaining are exclusively our own.
6. References


