Wh-indeterminates and Prosody in Hong Kong Cantonese

Yu-Yin Hsu[1], Anqi Xu[2]

[1]Hong Kong Polytechnic University, Hong Kong
[2] Speech, Hearing and Phonetic Sciences, University College London, UK
yyhsu@polyu.edu.hk, a.xu.17@ucl.ac.uk

Abstract

We report results of a speech production experiment about the prosody of *wh*-indeterminates and three types of sentences in Hong Kong Cantonese, and discuss our results in relation to the characteristics of focus prosody and the prosodic-syntactic effects on sentence final particles (SFPs). *Wh*-indeterminates refer to *wh*-phrases that are ambiguous between interrogative and indefinite readings. Chinese languages do not morphologically differentiate the meanings of *wh*-indeterminates, but they can be differentiated in some types of sentences marked by SFPs. In this study, we used statements as the baseline to systematically study the sentential prosody of *wh*- and *yes/no* questions as well as the constituent prosody expressed by *wh*-indeterminates therein. The results show that *wh*- and *yes/no* questions were distinguished from statements by the prosody of SFPs, and that the two readings of *wh*-indeterminates were distinguished in the regions of *wh*-phrases. We also found that *wh*-phrases and SFPs together formed a specific duration pattern that distinguishes questions from statements. Our results suggest that the speech prosodic organization considers and interacts with syntax-semantics.

Index Terms: *wh*-indeterminates, sentence final particles, focus prosody, syntax, Cantonese

1. Introduction

In this study, we examined the prosodic characteristics of a special lexical category – *wh*-indeterminates – which refers to *wh*-phrases that are ambiguous between interrogative and indefinite readings ([1] [2] [3]). For example, a *wh*-phrase in Cantonese like *matje* 'what' can be interpreted as a *wh*-interrogative 'what' in a *wh*-question (e.g., (1a)), or as an indefinite 'something' in a *yes/no* question (e.g., (1b)) or a statement (e.g., (1c)), although speakers often tend to use a reduced form (*mat/me* or *je*) in expressing the indefinite meaning 'something/anything' to avoid ambiguity.

(1) a. Zoengsaam maai-zo matje aa?
    Zoengsaam buy-ASP what WH-PARTICLE
    'What did Zoengsaam buy?'
  b. Zoengsaam maai-zo matje aa?
     Zoengsaam buy-ASP what Y/N-PARTICLE
    'Did Zoengsaam buy something?'
  c. Zoengsaam maai-zo (mat)je (gw)aa.
    Zoengsaam buy-ASP what S-PARTICLE
    'Zoengsaam probably bought something.'

Concerning Chinese languages, previous studies have identified that the occurrence of sentence final particles (henceforth SFPs) specifies the sentence types [4] [5] and functions and SFP is one of the contexts that licenses the indefinite reading of *wh*-phrases ([2] [3]). Therefore, Chinese *wh*-indeterminates not only are lexically ambiguous but are also

relevant to structural ambiguity [3] (cf. [6] for Mandarin), since SFPs are not obligatory, and sentence types and information structure are not often prosodically specified in Cantonese (e.g., [7]). Some phonology-syntax studies proposed that sentences are not necessarily distinguished by acoustic prominence, since Chinese languages use lexical tones and clausal types are already expressed by syntax [8] [9], while some studies reported that *wh*-questions generally had higher F0 contours than that of statements [10] [11].

Crosslinguistically, some studies have reported that *wh*-indeterminates while functioning as *wh*-interrogatives manifest more acoustic prominence than *wh*-indefinites in languages such as Korean [12], Japanese [13] [14] [15], and German [16]. Different results were reported for *wh*-indeterminates in Mandarin Chinese. Hu [17] reported that Mandarin speakers expressed *wh*-interrogatives acoustically different from *wh*-indefinites, 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 [18] reported two main findings; first, they reported some differences between *wh*-interrogatives and indefinites based on eight tokens (from a speech corpus [19]), but the syntactic and the phonetic contexts where these tokens occurred were different. Second, their own production study showed that participants did not acoustically disambiguate *wh*-interrogatives from indefinites. However, since all participants in their second study responded to the same two items for one context, it is difficult to draw a general conclusion for TwM. Hsu and Xu [6] reported that *wh*-interrogatives in TwM are acoustically distinguished from indefinites, in which *wh*-interrogatives' prosody is partially similar to what has been reported about focus prosody in Mandarin varieties ([20] [21] [22] [23] [24]), due mostly to the co-occurrence of the SFPs. In particular, [6] showed that *wh*-phrases and the SFPs form specific patterns that differentiate statements from questions.

To the best of our knowledge, no study has investigated whether prosody is used to differentiate *wh*-interrogatives from indefinites in Cantonese, and most prior studies about Cantonese focus prosody used experimental items without SFPs. Prior studies about Chinese focus prosody have examined the acoustic prominence on the answer to a *wh*-question (without SFP in the experimental items) (e.g., [21] [22]); few studies are about the *wh*-phrases themselves that induce focus interpretations [25] [6]. Moreover, while different acoustic devices are used to express focus, it has been consistently reported for Chinese languages that focus units show on-focus F0 rising and lengthening [22] [23] [24].

The phenomenon of *wh*-indetermintates provides us a new context to systematically examine linguistic theoretical proposals and experimental acoustic findings, and to explore whether sentential prosodic organization interacts with focus

marking in this new syntactic context (cf. [24]). Therefore, we used a speech production experiment to study whether and how native speakers of Hong Kong Cantonese (HKC) prosodically differentiate *wh*-interrogative from the indefinite readings of wh-phrases, and whether and how HKC speakers prosodically distinguish focus units and sentence types.

2. Method

2.1. Stimuli

We examined the prosody of *wh*-questions and yes/no questions containing a *wh*-indeterminate by using statements as the baseline, as shown in Table 1. Target *wh*-questions and statements were with 8-9 syllables in the same tonal format, i.e., a monosyllabic subject pronoun in Tone5 (i.e., 你 *nei5* 'you' and 我 *ngo5* 'I') followed by a disyllabic modal auxiliary 可以 *ho2yi5* 'can', and a disyllabic helping verb 幫手 *bong1sau2* 'help'. Next, the main verb immediately before the target *wh*-phrase was a Tone3 monosyllabic verb (e.g., 教 *gaau3* 'teach', and 借 *je3* 'borrow'). The tonal environment of *yes/no* questions' is identical to that of *wh*-questions and statements, except that the auxiliary was formed in the A-not-A form (可唔可以 *ho2-m4-ho2-yi5*). These choices were made due to our three Cantonese consultants' suggestions and due to the lexicon limitation of the verbs while we considered the need of having enough parallel stimuli across sentence types and maintaining the same tonal contexts that hosted the *wh*-indeterminates. Right after the main verb was the *wh*-indeterminate target and it was immediately followed by a Tone3 monosyllabic SFP (i.e., *aa3* for questions, and *gwaa3* for statements).

Table 1: *Examples of target sentences with 'what'.*

Sentence Types	Format of Target Sentences					
	T5 subj	T2-T5 'can'	T1-T2 'help'	T3 V	Wh	T3 SFP
Wh-Q	你	可以	幫手	教	乜嘢	啊?
	'What can you help to teach?'					
Yes/no Q	你	可唔可以	幫手	教	乜嘢	啊?
	'Can you help to teach something?'					
Statement	我	可以	幫手	教	乜嘢	㗎。
	'I probably can help to teach something.'					

Four *wh*-phrases of different lengths were used (one monosyllabic *me1* 'what', and three disyllabic *wh*-phrases: *mat1ye5* 'what', *bin1dou6* 'where', and *bin1go3* 'who/which one') to construct each of the three types of sentences by using five versions of Tone3 verbs. In total, 60 target sentences (3 sentence types x 4 *wh*-phrases x 5 verbs) and 40 filler sentences (in different structure with no SPFs) were used in an experimental session. As shown in (2)-(4), each target trial consisted of a pre-recorded leading context (A) (18 character long), and a target sentence (B) that participants used to respond. The leading contexts were pre-recorded by a female speaker of Cantonese born and raised in-Hong Kong.

(2) *Wh*-Q
A: 呢個係咪廣告比賽啊？我識影相，拍片都識㗎。
'Is this an advertising competition? I know photography and filming.'
B: 你可以幫手拍乜嘢啊？ 'What can you help to film?'

(3) *Yes/no* Q
A: (Same leading context as the one in the *wh*-question)
B: 你可唔可以幫手拍嘢啊？ 'Can you film something?'
(4) Statement
A: 我要拍幾條片出嚟啊。有邊個可以幫下手啊？
'I plan to make some videos. Is there anyone who can help?'
B: 我可以幫手拍嘢㗎。
'I probably can help [you] to film something.'

2.2. Participants

10 female native Cantonese speakers born and raised in Hong Kong (mean age 21.2 (± 1.03 SD) years), who were university students, joined our study. None reported any reading-hearing problems. Participant were paid HK$50 after the experiment.

2.3. Procedure

Each participant first signed an informed consent form and filled out a background questionnaire. The experiment was conducted in Hong Kong in a sound-attenuated speech lab with a Focusrite Scarlett 2i2 sound interface, and a Telefunken M-80 dynamic microphone that was calibrated to measure intensity. Each participant was 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, so that no similar target item occurred immediately adjacent. Participants were asked to listen to the leading context, and then read the target sentence aloud as casually and naturally as possible. No instructions were given regarding focus or emphasis. Participants produced each sentence twice, and additional repetitions were only allowed in cases of mispronunciation or hesitation. Their speech was recorded in the .wav format at a sampling rate of 44.1 kHz with 16-bit quantization. Three practice trials were before the main trials. Each experiment lasted about 25 to 30 minutes.

2.4. Acoustic measurements

The acoustic measurements were generated by ProsodyPro 5.7.6 [26] for duration, intensity, and time normalized fundamental frequency (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 creaky voice and pitch halving or doubling. Linear Mixed Effects models were conducted on the duration and the intensity of SFPs and each syllable of *wh*-phrases in all items using the lme4 package [27] in R [28]. The model first included random intercepts for item and speaker and also by-speaker, by-item, by-speaker-item-interaction random slopes for the sentence type (statement, *wh*- and *yes/no* questions). Sentence type was then added as a potential fixed effect. The significance of the main effect was evaluated by likelihood ratio tests.

F0 was time-normalized across tokens by dividing each syllable into 10 intervals equal in time and the trimmed F0 values were calculated. To observe the F0 realization during production, the speed of fundamental frequency shift, i.e., F0 velocity (in semitone/s) was also measured. The graphical analysis of F0 and F0 velocity were performed by Smoothing Spline ANOVA (SS-ANOVA) [29]. 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. Only the F0 (velocity) curves of statements and *wh*-questions were reported, because the items of *yes/no*-questions had more syllables.

3. Results

3.1. F0 contours

Figure 1 shows SS-ANOVA statistic-estimates (central solid and dashed lines) and confidence intervals (color ribbons) for F0 of statements and *wh*-questions containing four different *wh*-phrases ("wh", i.e., the monosyllabic 'what1' and three disyllabic *wh*-phrases: 'what2', 'where', and 'who'). The overall F0 patterns across four *wh*-phrases showed that statements were prosodically distinguished from questions. In the sentence-initial position, the statements with 'where' had marginally higher F0 than *wh*-questions, while little divergences were found for sentences with the other three *wh*-phrases. In the sentence-medial position, only *wh*-questions with 'what2' showed marginally higher F0 at the regions of the auxiliary 'can' and the adverb.
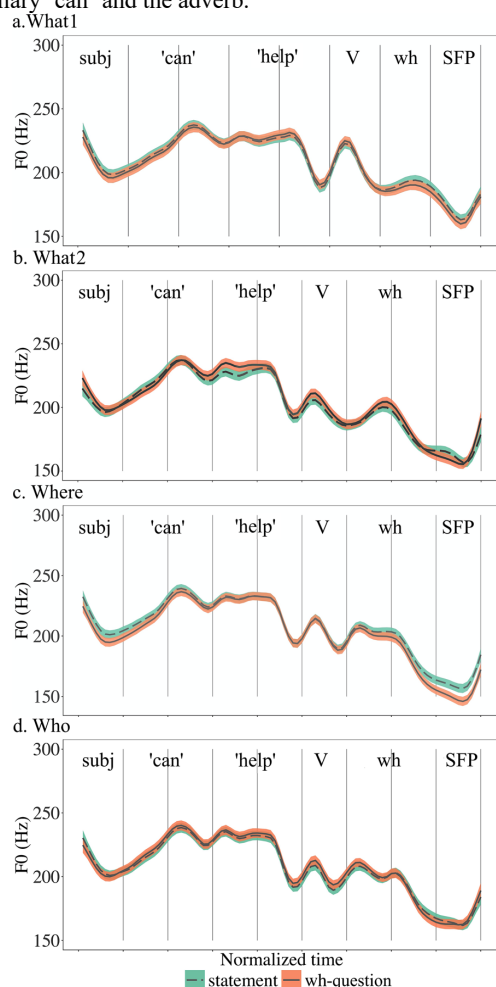

Figure 1: *SS-ANOVA results of time-normalized F0 of statements and wh-questions with different wh-phrase. Vertical lines indicate syllable boundaries.*

In the regions of "wh" and "SFP", statements with 'where' was significantly higher than that in *wh*-questions, and marginally higher F0 was found in statements with 'what1' and 'what2. The lenghth of *wh*-phrases (monosyllabic vs. disyllabic) did not show effects on F0 (velocity) patterns. The

transition of F0 velocity was the most prominent in items with 'what2' and marginally with 'who', as shown in Figure 2.
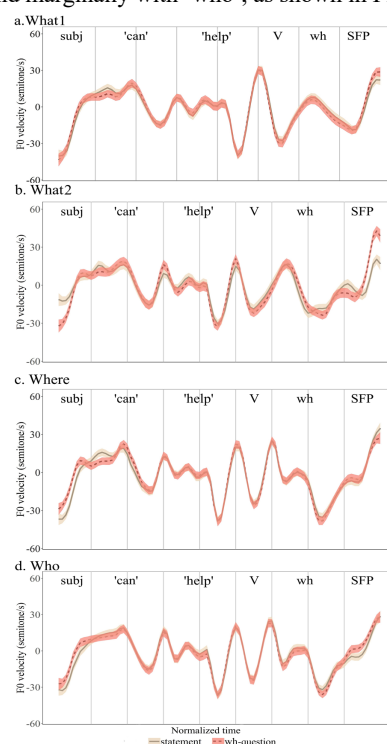

Figure 2: *SS-ANOVA results of time-normalized F0 velocity contours of statements and wh-questions. Vertical lines indicate syllable boundaries.*

3.2. Duration

Among items with 'what1', the main effect of sentence type was significant in the region of the SFP ($\chi^2 = 8.504$, df = 2, $p = .014$) but not in the region of the *wh*-phrases ($\chi^2 = 4.022$, df = 2, $p = .134$). The duration of the SFP was longer in statements than *wh*-questions ($p = .015$) and *yes/no*-questions ($p = .021$), while the two questions did not differ significantly ($p = .980$).

Among items with 'what2', the sentence type did not significantly inflauence the duration of the first syllable of the *wh*-phrases ($\chi^2 = 2.853$, df = 2, $p = .240$), while the effect was significant on their second syllable ($\chi^2 = 14.149$, df = 2, $p < .001$), i.e., the second syllable was longer in *wh*-questions ($p = .025$) and *yes/no*-questions ($p < .001$) than the one in the statements. Sentence type also significantly affected the duration of SFP ($\chi^2 = 13.191$, df = 2, $p = .001$). SFPs were longer in statements than in *yes/no*-questions ($p < .001$).

When the *wh*-phrase was 'where' (Figure 1), the F0 cues were evident in the region of the *wh*-phrase and the SFP. Here, similar tendency was observed in syllable duration. The main effect of sentence type was significant on both the first ($\chi^2 = 13.215$, df = 2, $p = .001$) and the second ($\chi^2 = 9.882$, df = 2, $p = .007$) syllable of 'where'. The duration of the first syllable of 'where' in *wh*-questions was longer than that in both *yes/no*-questions ($p = .005$) and statements ($p < .001$). Its second syllable was longer in *wh*-question ($p < .001$) and *yes/no*-questions ($p < .001$) than that in statements. Sentence types influenced the duration of the SFP ($\chi^2 = 10.501$, df = 2, $p = .005$). The SFP in statements had significantly longer duration than *wh*-questions ($p = .002$) and *yes/no*-questions ($p < .001$).

For sentences with 'who', the main effect of sentence types was likewise significant in both the first ($\chi 2 = 8.848$, df = 2, $p = .012$) and the second syllable ($\chi 2 = 15.736$, df = 2, $p < .001$) of the *wh*-phrase. The duration of the first syllable of 'who' in *yes/no*-questions was longer than the one in the statements ($p = .006$) and *wh*-questions ($p = .036$). The second syllable of 'who' was longer in the *wh*-questions ($p < .001$) and *yes/no*-questions ($p < .001$) than the statements. Moreover, the main effect of sentence type was significant on the duration of the SFP ($\chi 2 = 14.921$, df = 2, $p < .001$). The SFP in statements was longer than the one in *yes/no*-questions ($p < .001$). The results of duration across four types of *wh*-words showed that focus prosody interacted with the organization of sentence prosody: On-focus lengthening and the shortening of question-SFPs were found to distinguish questions from statements.
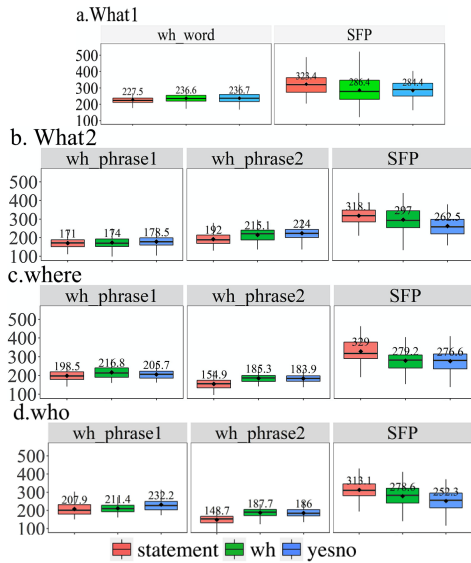

Figure 3: *Boxplots for the duration (in ms) of wh-phrases in statements, wh- and yes/no-questions.*

3.3. Intensity

Table 2 shows results of linear mixed models. Unlike the duration, the intensity did not differentiate sentence types or different readings of *wh*-phrases (see the boxplots in Figure 4).

Table 2: *Results of LMM. Wh-s1 and wh-s2 indicate the 1st syllable and the 2nd syllable of the wh-phrase.*

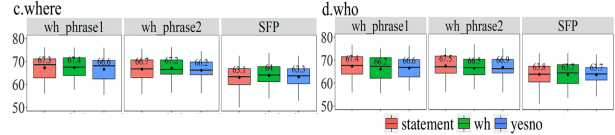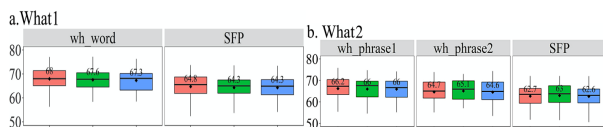|	X2	df	*p*		$\chi 2$	df	*p*
What1				**What2**			
wh-word	3.24	2	.20	wh-s1	.24	2	.89
SFP	1.49	2	.48	wh-s2	1.36	2	.51
NA				SFP	.32	2	.85
Where				**Who**			
wh-s1	2.47	2	.29	wh-s1	5.29	2	.07
wh-s2	2.32	2	.31	wh-s2	2.71	2	.26
SFP	3.61	2	.165	SFP	.075	2	.96

Figure 4: *Boxplots for the intensity (in dB) of wh-phrases in statements, wh- and yes/no-questions.*

4. Discussion

We explored whether Hong Kong Cantonese speakers prosodically distinguish *wh*-focus from indefinite readings, and whether the organization of sentential prosody interacts with focus marking. Our results show some support to these inquires. Concerning acoustic measurements, intensity did not seem to be used to distinguish *wh*-interrogatives from the indefinite readings, nor did intensity signal the sentence types. The length of *wh*-phrases also did not show influences. While no significant differences of F0 were found across most of the syllables in a sentence, SFPs of *wh*-questions showed lower F0 than SFPs of statements but the effect was only significant when the *wh*-phrase was 'where'. The speed shift of F0 was also more obvious in the regions of SFPs (Figure 2). In addition to F0, our results show an interesting and consistent pattern of duration modulated across four *wh*-phrases.

Duration was used to distinguish *wh*-interrogatives from indefinite readings. However, unlike the on-focus lengthening effect reported in previous studies, in our study, only the 2nd syllable of the disyllabic *wh*-phrases in questions was significantly longer than that one in statements, while the 1st syllable was lengthened in items of 'where' and 'who', but not in items of 'what2'; no lengthening effects were found with the monosyllabic 'what1', despite the fact that the 1st syllables of these *wh*-phrases are all Tone1. This might seem rather surprising until we looked at the results of the SFPs (which immediately followed the *wh*-phrases), that is, SFPs in both questions were all significantly shorter than the SFPs in statements. Considering the duration patterns of *wh*-phrases and SFPs together, we think that these two phrases may cooperate in marking focus prosody in Cantonese, i.e., the occurrence of SFPs explicitly defines the sentence types, and while there is a need to maintain lexical tones, the information indicated by SFPs requires the prosodic organization to comply with focus marking, as it was shown by the acoustic differences of the Tone3 SFPs between statements and questions. This influence from SFPs may be a reason why not every syllable of the *wh*-interrogatives were lengthened as reported in other studies.

In sum, our results suggest that the internal organization at different structural levels (syntax and information structure) may interact with the prosody organization. We would expect to see similar inter-level interactions in other tone languages in future studies. The results will advance our understanding further about the human language system.

5. Acknowledgements

6. References

[1] S.-Y. Kuroda, "Generative grammatical studies in the Japanese language," MIT, Cambridge, MA, 1965.

[2] M. Haspelmath, Indefinite Pronouns, Oxord: Clarendon Press, 1997.

[3] Y.-h. A. Li, "Indefinite Wh in Mandarin Chinese," *Journal of East Asian Linguistics,* vol. 1, no. 2, pp. 125-155, 1992.

[4] C. N. Li and S. A. Thompson, Mandarin Chinese: a Functional Reference Grammar, Berkeley: University of California Press, 1981.

[5] W. L. Wu, "Sentence-final particles in Hong Kong Cantonese: Are they tonal or intonational?," in *INTERSPEECH 2009, the 10th Annual Conference of the International Speech Communication Association*, Brighton, 2009.

[6] Y.-y. Hsu and A. Xu, "Sentence Prosody and Wh-Indeterminates in Taiwan Mandarin," in *INTERSPEECH 2019 the 20th Annual Conference of the International Speech Communication Association*, Graz, Austria, 2019.

[7] Y.-y. Hsu, A. Xu and H. Ngai, "Focus Prosody in Cantonese and Teochew Noun Phrases," in *9th International Conference on Speech Prosody 2018*, Poznań, Poland, 2018.

[8] L. Xu, "Manifestation of informational focus," *Lingua,* vol. 114, pp. 277-299, 2004.

[9] K. Li, "An OT Analysis of Informational Focus in Mandarin Chinese," in *Proceedings of the 21st North American Conference on Chinese Linguistics*, Smithfield, R, 2009.

[10] X.-n. S. Shen, The prosody of Mandarin Chinese, Berkeley, Los Angeles, Oxford: University of California Press, 1990.

[11] L. L.-s. Cheng, "What wh-questions tell us about the syntax-prosody mapping," in *International Roundtable on Current Issures in Chinese Prosodic Syntax*, Hong, 2017.

[12] S.-a. Jun and O. Mira, "A prosodic analysis of three types of wh-phrases in Korean," *Language and Speech,* vol. 39, no. 1, pp. 37-61, 1996.

[13] S. Ishihara, "Intonation and interface conditions," MIT, Cambridge, MA, 2003.

[14] M. Deguchi and Y. Kitagawa, "Prosody and wh-questions," in *In North Eastern Linguistic Society (NELS) 32, ed. by Mako Hirotani*, Amherst, MA, University of Massachusetts, Graduate Linguistic Student Association., 2002, pp. 73-92.

[15] Y. Kitagawa, "When We Fail to Question in Japanese," in *Proceedings of the 2nd workshop on prosody, syntax, and information structure (WPSI 2)*, 2007.

[16] H. Truckenbrodt, "An analysis of prosodic F-effects in interrogatives: Prosody, syntax and semantics," *Lingua,* vol. 124, pp. 131-175, 2013.

[17] F. Hu, "A prosodic analysis of wh-words in Standard Chinese," in *International Conference on Speech Prosody 2002*, Aix-en-Provence,France, 2002.

[18] S.-i. Shyu and T.-l. Tung, "Taiwan Mandarin Wh-phrases and Prosody," *Studies in Prosodic Grammar,* Accepted.

[19] Y.-F. Wang, J.-g. Chen, D. Treanor and H.-M. Hsu, "Exclusivity, contingency, exceptionality and (un)desirability: A corpus-based study of Chinese chufei ('unless') in spoken and written discourse," *Language & Communication,* vol. 37, pp. 40-59, 2014.

[20] S. Jin, "An acoustic study of sentence stress in Mandarin Chinese," in *Columbus, OH*, Ohio State University, 1996.

[21] Y.-y. Hsu and A. Xu, "Focus Acoustics in Mandarin Nominals," in *INTERSPEECH 2017 the 18th Annual Conference of the International Speech Communication Association*, Stockholm, Sweden, 2017.

[22] S.-W. Chen, B. Wang and Y. Xu, "Closely related languages, different ways of realizing focus," in *Interspeech*, Brighton, UK, 2009.

[23] Y. Chen and C. Gussenhoven, "Emphasis and tonal implementation in Standard Chinese," *Journal of Phonetics,* vol. 36, no. 4, pp. 724-746, 2008.

[24] Y.-y. Hsu and J. S. German, "Prosodic Organization and Focus Realization in Taiwan Mandarin," in *the 32nd Pacific Asia Conference on Language, Inforamtion and Computation (PACLIC)*, Hong Kong, 2018.

[25] S. Feng, Hanyu Yunlü Jufaxue [Mandarin Prosodic Syntax], Shanghai, China: Shangwu Publishing Co., 2013.

[26] Y. Xu, "A Tool for Large-scale Systematic Prosody Analysis," in *TRASP 2013*, Aix-en-Provence, France, 2013.

[27] D. Bates, M. Mächler, B. Bolker and S. Walker, "Fitting Linear Mixed-Effects Models Using lme4," *Journal of Statistical Software,* vol. 67, no. 1, pp. 1-48, 2015.

[28] R Core Team, "R: A language and environment for statistical," R Foundation for Statistical Computing, 2013. [Online]. Available: http://www.R-project.org/.

[29] C. Gu, Smoothing Spline ANOVA Models, New York: Springer, 2013.