Interaction of prosody and syntax-semantics in Mandarin *wh*-indeterminates

Yu-Yin Hsu, and Anqi Xu

ARTICLES YOU MAY BE INTERESTED IN

Interaction of prosody and syntax-semantics in Mandarin *wh*-indeterminates

Yu-Yin Hsu[1,a)] **and Anqi Xu**[2]

[1]*The Hong Kong Polytechnic University, Hung Hom, Hong Kong, China*
[2]*University College London, London, United Kingdom*
yu-yin.hsu@polyu.edu.hk, a.xu.17@ucl.ac.uk

Abstract: This paper reports on two speech-production experiments focused on Putonghua and Taiwan Mandarin sentence-final particles and *wh*-phrases that have interrogative or indefinite readings in three contexts: *yes/no* questions, *wh*-questions, and statements. Sentence-final particles were found to influence focus-prosody through right-edge shortening and lowering of F0 and intensity of *wh*-phrases, thus distinguishing *wh*-interrogatives from indefinites and questions from statements. Speakers adopt multidimensional acoustic strategies to shape intonation: while maintaining the lexical tones, prosody interacts with the organization imposed by syntax, semantics, and focus. The two varieties of Mandarin differ in the extent to which their prosodic differences represent such syntactic-semantic information.
© 2020 Acoustical Society of America

1. Introduction

Linguists have a long-standing interest in how tone languages prosodically resolve linguistic ambiguity and signal sentence types. This is a challenging research topic as acoustic variations cross-linguistically associated with highlighting information are simultaneously used to distinguish word meaning. The experimental study presented in this paper contributes to this line of inquiry by examining the prosodic nature of three types of sentences in Mandarin and a special lexical category, known as *wh*-indeterminates. *Wh*-indeterminates are *wh*-phrases that can have either an interrogative or indefinite reading.[1] For example, the Mandarin *wh*-phrase *shenme* can be interpreted as the *wh*-interrogative "what" *in situ* in a *wh*-question, or as an indefinite noun "something," in a *yes/no* question (Table 1). Previous studies of syntax have concluded that the occurrence of the sentence-final particle (SFP), *ma* for *yes/no* questions in Mandarin, licenses an indefinite reading of *wh*-phrases.[2] In statements with the SFP *ba*, indicating weak epistemic judgment,[3] *shenme* also expresses an indefinite reading. Moreover, because SFPs are not obligatory, Mandarin *wh*-indeterminates are not only lexically ambiguous, but syntactically ambiguous, too. This study uses identically constructed sentential tonal contexts to identify the interactive roles of prosody, semantics, syntax, and information structure in speech production.

Regarding prosody as a disambiguation device, a number of prior studies of Korean,[4] Japanese,[5,6] and German[7] report that *wh*-phrases are more acoustically prominent when they function as *wh*-interrogatives, compared to when they function as indefinites. However, some scholars could not identify any acoustic differences that could be used to distinguish between those two readings.[8] According to Hu,[9] speakers express *wh*-interrogatives in Putonghua (henceforth ChM) with a mean F0 higher than when they produce indefinites. However, Shyu and Tung[10] reported that speakers did not prosodically disambiguate *wh*-phrases in Taiwan Mandarin (henceforth TwM). In short, the findings from research about Mandarin *wh*-indeterminates' prosody remains contradictory and inconclusive, possibly due to subtle differences between the different variations of Mandarin. This study aims to clarify which mechanisms are at play.

Systematic investigation of the influence of prosody on sentence meaning remains limited. A number of existing studies have concluded that Chinese does not use prosody to signal sentence information because Chinese languages use lexical tones and express information about clausal types through syntax.[11] However, other studies found that *wh*-questions had a higher overall F0 when compared to statements.[12] To the best of our knowledge, few acoustic studies of focus prosody have closely examined the *wh*-phrases themselves.[13] Instead, most have focused on the acoustic prominence of answers to Chinese *wh*-questions with no SFP.[14,15] Prosodic correlates of focus are well-documented for ChM, and different patterns of focus realization have been

a)Author to whom correspondence should be addressed, ORCID: 0000-0003-4087-4995.

Table 1. Example sentences including *shenme* "what/something."

Sentence Type	T3 subj.	T2-T3 "can"	T1-T2 "help"	T4 V	*wh*-phrase	T1 SFP	English Translation
a. *Wh*-Q	你 "you"	可以	幫忙	帶 "bring"	什麼 "what"	呢 "*ne*"	**What** can you help bring [to me]?
b. *Yes/no* Q	你 "you"	可以	幫忙	帶 "bring"	什麼 "something"	嗎 "*ma*"	Can you help to bring [me] **something**?
c. Statement	我 "I"	可以	幫忙	帶 "bring"	什麼 "something"	吧 "*ba*"	I can help to bring [you] **something**.

reported for different varieties of Mandarin.[16] For example, it is generally assumed that post-focal compression of pitch range will be found after focalized simple nouns[17] and even across morphosyntactic boundaries[18] in ChM, but that no such effects are found among speakers of TwM, with the exception of young people around 20 years of age.[19] Several Chinese languages exhibit on-focus F0 rising and lengthening of whole focus phrases.[20,21] Yet, little research has hitherto examined either the acoustics of SFPs[22] or their impacts on focus and/or sentence intonation.[23]

Wh-indeterminates are therefore an important, yet under-studied, phenomenon through which we can extend our knowledge base on focus prosody and syntax-semantics by focusing on how they are integrated by native speakers. To achieve this, we designed two speech-production experiments to ascertain whether the previously reported focus-prosody similarities between speakers of two varieties of Mandarin, ChM and TwM, could also be observed in their prosody's interactions with other linguistic levels. All SFPs in our study bore tone 1, and the target *wh*-phrases were embedded in nearly identical (supra)segmental carrier sentences. From a purely phonotactic and/or syntactic perspective, no differences in acoustic patterns were expected across the different types of sentence. Regarding focus prosody, each syllable in a *wh*-phrase within a *wh*-question was expected to rise in F0, intensity, and length. This is because the *wh*-phrase is expected to be the focus of all *wh*-questions. In contrast, the F0, intensity, and length of the *wh*-SFP in such questions was expected to decrease as a result of post-focus compression in Mandarin. However, if linguistic levels *other than* information structure (focus) interact with prosody, consistent patterns differing from those mentioned above were expected to be observed. If *wh*-phrases used as interrogatives are distinguished prosodically from those indefinites, this would suggest that prosodic organization interacts with semantics. Likewise, on the sentential level, the overall prosody of *yes/no* questions was expected to differ from both constituent foci (*wh*-questions) and statements (non-focus), owing to the status of polar proposition focus. If both *yes/no* and *wh*-questions' prosody tend to be similar—especially on the SFPs—this would suggest that prosodic organization not only interacts with syntax (clause-typing), but also focus prosody, as such a pattern would not be in keeping with expected post-focus compression in Mandarin.

2. Methods

2.1 Participants

We recruited 20 female native Mandarin speaking university students, 10 of whom were born and raised in northern provinces of China (hereafter, the ChM group; mean age $\pm$ SD: 24 $\pm$ 3.32 years), and the other 10 having been raised in Taiwan (TwM group; mean age $\pm$ SD: 20.1 $\pm$ 0.83 years). None reported having any history of speaking or hearing problems. All ChM participants reported native fluency in Mandarin on a seven-point self-report scale (i.e., assigned themselves scores of 7) and reported non-fluency in other Chinese languages (scores of 1–3). All participants in the TwM group also reported native fluency in Mandarin. Four reported non-fluency in Taiwanese Min (scores of 1–3), and six reported intermediate fluency in Taiwanese Min (scores of 4–5).

2.2 Stimuli

All target sentences were constructed in the same tonal format (see Table 1). Three *wh*-phrases (*shen2.me* "what," *na3.li3* "where," and *shui2* "who") and five tone 4 verbs were used to construct experimental sentences, all of which were checked for naturalness. Despite the limitations of the lexicon, we were able to construct sentences appropriate for fulfilling our research needs: (1) sufficient parallel stimuli across sentence types, (2) to be able to examine the potential effects of *wh*-word lengths, and (3) assurance that all target sentences had the same tonal contexts. In each experimental session, the study used 45 target sentences (3 sentence types $\times$ 3 *wh*-phrases $\times$ 5 verbs), with an additional 40 filler sentences that featured other syntactic and tonal structures, not including any SFPs. Three examples of target sentences are shown below.

(1) *Wh*-Q → A: "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? "What can you help bring [to me]?"

(2) *Yes/no* Q →A: [as (1A)] B: 你可以幫忙帶什麼嗎? "Can you help bring something [to me]?"

(3) Statement → A: "This Saturday, we are going on a picnic by the big lake. Can anyone help to bring stuff for the picnic?" B: 我可以幫忙帶什麼吧。. "I can help to bring [you] something."

3. Procedure

The experiments were conducted in a sound-attenuated speech lab, with a dynamic microphone calibrated to measure intensity. After signing a consent form and completing a demographic survey, each participant was seated in front of a computer screen while wearing headphones. Each participant was then asked to listen to a pre-recorded 25-character context and was then asked to read the target sentence aloud twice, as casually and as naturally as possible. Each session involved three practice trials. Repetitions were permitted during the main experiment in cases of mispronunciation or hesitation. Target sentences were presented on screen, either in traditional (TwM group) or simplified (ChM group) characters. These remained visible until the participant had finished reading. Trials were pseudo-randomized to ensure that no items with the same words in different sentence types or with different words but the same sentence type, occurred adjacent to each other. The recordings were in .wav format at a sampling rate of 44.1 kHz with 16-bit quantization. Each experiment lasted about 30 min. Participants received 60 HKD (about 8 USD) as an incentive to participation.

4. Data analysis

Measurements were generated by PROSODYPRO.[24] Syllable boundaries were determined using both visual and auditory information. Vocal pulses were manually corrected in cases of pitch halving/doubling or in cases of a creaky voice. Time-normalized F0 was analyzed by smoothing spline analysis of variance (SS-ANOVA)[25] for sentential prosody. Linear mixed-effects models of the duration and intensity of each syllable of *wh*-phrases and SFPs were fitted.[26] Each model first included random effects, ITEM and SPEAKER, with slopes. SENTENCE TYPE, LANGUAGE, and their interactions were added as fixed effects and evaluated using a likelihood ratio. Tukey's tests were conducted *post hoc*.

5. Results

5.1 Sentential F0 contours

The study found that while the lexical tones were maintained, both ChM and TwM speakers clearly distinguished statements from questions in sentence-initial positions, *wh*-phrases, and SFPs (see Fig. 1). Monosyllabic and disyllabic *wh*-phrases patterned similarly to each other, and all SFPs exhibited F0 raising. Nevertheless, the productions of the ChM and TwM groups differed, with the former tending to raise the F0 of syllables in *wh*-phrases and the SFPs in *wh*-questions (1) more often than in statements and (2) dramatically more than for *yes/no* questions. In contrast, the TwM speakers tended to use F0 variation only to distinguish questions from statements.

5.2 Duration and intensity of wh-indeterminates and sentence-final particles

5.2.1 *wh*-indeterminates

In both groups, SENTENCE TYPE significantly influenced the duration and intensity of *wh*-indeterminates (Table 2). Monosyllabic "who" was lengthened and spoken with higher intensity
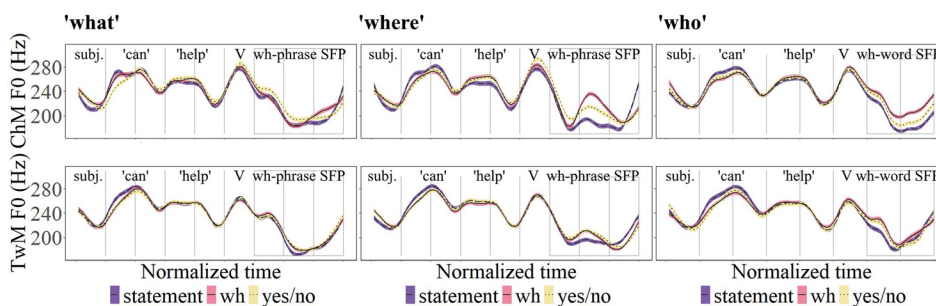

Fig. 1. (Color online) SS-ANOVA plots of F0 of sentence types in Putonghua (ChM, top row) and Taiwan Mandarin (TwM, bottom row). Lines represent the means, and ribbons around the lines show 95% confidence intervals. Non-overlapping ribbons indicate statistically significant differences.

Table 2. Effect of SENTENCE TYPE on duration and intensity of each syllable of *wh*-phrases. Dur = duration; Int = intensity; state = statement.

"What"	shen2	me
Dur	$\chi^2 = 8.316$, df $= 2$, $p = 0.016^a$ (*wh* > state & *yes/no*)	$\chi^2 = 19.570$, df $= 2$, $p < 0.001^a$ (state > *wh* & *yes/no*)
Int	$\chi^2 = 9.680$, df $= 2$, $p = 0.008^a$ (*wh* & *yes/no* > state)	$\chi^2 = 18.959$, df $= 2$, $p < 0.001^a$ (*yes/no* > *wh* > state)
"Where"	na3	li3
Dur	$\chi^2 = 13.620$, df $= 2$, $p = 0.001^a$ (*wh* > state & *yes/no*)	$\chi^2 = 26.927$, df $= 2$, $p < 0.001^a$ (state > *wh* > *yes/no*)
Int	$\chi^2 = 11.409$, df $= 2$, $p = 0.003^a$ (*wh* & *yes/no* > state)	$\chi^2 = 25.273$, df $= 2$, $p < 0.001^a$ (*wh* & *yes/no* > state)
"Who"	shui2	
Dur	$\chi^2 = 19.220$, df $= 2$, $p < 0.001^a$ (*wh* & state > *yes/no*)	
Int	$\chi^2 = 20.832$, df $= 2$, $p < 0.001^a$ (*wh* & *yes/no* > state)	

[a]Statistically significant.

in *wh*-questions. Disyllabic "what" and "where" exhibited some edge effects, as reported in previous studies for *wh*-focus: i.e., the initial lengthening of *wh*-focus significantly distinguished them from indefinites. Yet, in contrast to previously reported findings regarding on-focus lengthening, the final syllables of disyllabic *wh*-focus were significantly shorter in *wh*-questions than for statements, with their intensity being higher than for that used in statements.

      Speakers of both Mandarin varieties tended to voice the stimuli similarly, except for the word "who" in *wh*-questions (Fig. 2), where TwM speakers had both a longer duration ($\chi^2 = 4.442$, df $= 1$, $p = 0.035$) and a higher intensity ($\chi^2 = 6.040$, df $= 1$, $p = 0.014$) compared to ChM speakers. Moreover, despite the many similarities between the productions of the two participant groups, we identified a significant interactive effect of LANGUAGE and SENTENCE TYPE on duration ($\chi^2 = 12.224$, df $= 3$, $p = 0.007$) and intensity ($\chi^2 = 10.232$, df $= 3$, $p = 0.017$) of the first syllable of "what" (i.e., *shen2*). That is, while TwM members did not vary this syllable significantly across sentence types, ChM members took longer to say "what" and said it with higher intensity in *wh*-questions, compared to *yes/no* questions or statements ($p < 0.001$). In addition, the above-mentioned interaction effect had a significant impact on the duration of monosyllabic "who" ($\chi^2 = 9.789$, df $= 2$, $p = 0.007$). In the ChM data, the duration of the *wh*-questions was greater than that of the *yes/no* questions ($p < 0.001$), but no such difference was found in the TwM data ($p = 1.000$). However, the contrast between the *wh*-questions and statements was significant in TwM ($p = 0.004$) but not in ChM ($p = 1.000$). In sum, ChM speakers tended to vocally distinguish more clearly between the two question types, with TwM speakers distinguishing only between questions and statements.

5.2.2 Sentence-final particles

SENTENCE TYPE was also significantly correlated with both the duration and the intensity of SFPs. That is, across both language groups, the SFPs of *yes/no* questions were significantly longer and greater in intensity than for statements. *Wh*-questions showed the same pattern, except in the duration of SFPs following "who" (Table 3).

      As Fig. 2 shows, the SFPs of questions were significantly longer and higher in intensity than those of statements for speakers in both groups. However, the two groups signaled
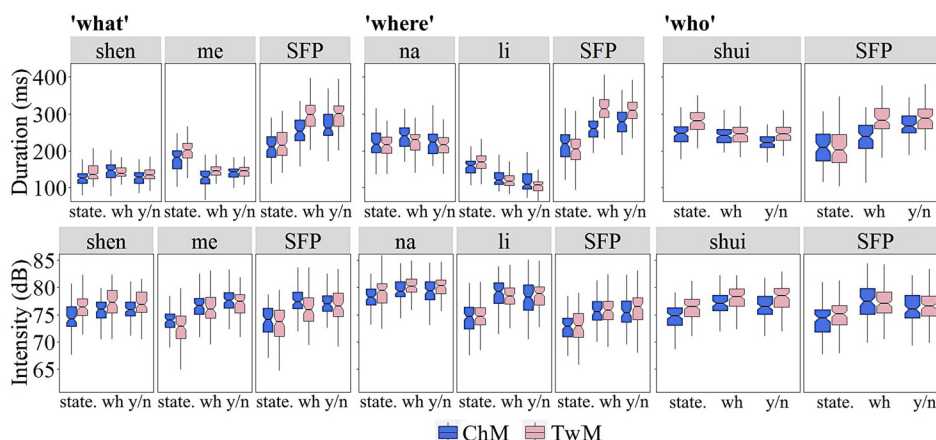

Fig. 2. (Color online) Boxplots of syllable duration and intensity of *wh*-phrases and SFPs, by language group.

Table 3. Effect of SENTENCE TYPE on duration and intensity of SFPs.

Preceding-*wh*	"What"	"Where"	"Who"
Duration (Post-hoc)	$\chi^2 = 17.989$, df $= 2$, $p < 0.001$[a]	$\chi^2 = 26.400$, df $= 2$, $p < 0.001$[a]	$\chi^2 = 20.843$, df $= 2$, $p < 0.001$[a]
	(*wh* & *yes/no* > statement)	(*yes/no* > *wh* statement)	(*yes/no* > *wh* & statement)
Intensity (Post-hoc)	$\chi^2 = 20.204$, df $= 2$, $p < 0.001$[a]	$\chi^2 = 17.317$, df $= 2$, $p < 0.001$[a]	$\chi^2 = 16.366$, df $= 2$, $p < 0.001$[a]
	(*wh* & *yes/no* > statement)	(*wh* & *yes/no* > statement)	(*wh* & *yes/no* > statement)

[a]Statistically significant.

SENTENCE TYPE differently in terms of duration of SFPs in "what" ($\chi^2 = 9.160$, df $= 3$, $p = 0.027$), "where" ($\chi^2 = 26.014$, df $= 3$, $p < 0.001$), and "who" ($\chi^2 = 20.843$, df $= 2$, $p < 0.001$) conditions. In the "who" condition, the TwM group took significantly longer to utter *wh*-questions SFPs than to utter statement SFPSs ($p < 0.001$), whereas the ChM group did not ($p = 0.419$). A significant difference in the SFPs for *yes/no* questions and those for *wh*-questions was found only in the ChM data ($p = 0.002$).

6. Discussion

This study aimed to establish whether and how speakers of two varieties of Mandarin prosodically distinguish between two interpretations of *wh*-indeterminates and between sentence types. Our results show that, for both groups of speakers, the sentential-prosody mechanism not only takes into account lexical tones, but also sentence types and focus-marking. Word length also seems to influence focus-marking. Our target sentences across sentence types bore identical lexical tones, yet consistent patterns were found that distinguished statements from questions. We also observed that specific acoustic patterns of *wh*-focus were affected by SFPs.

First, at the word level, while each syllable needed to reflect lexical tonal information through F0 (i.e., similar patterns of F0 contours that we observed across various contextual placements of each syllable), F0 heights were nevertheless used to signal *wh*-focus, semantics, and sentence types. Consistent with findings from previous focus studies, all types of *wh*-phrases in our study exhibited higher F0 in *wh*-questions than in statements and across both language groups. This indicates that speakers of both varieties of Mandarin acoustically distinguish *wh*-focus (in *wh*-questions) from indefinites (in statements). Moreover, while the overall F0 contours of *yes/no* and *wh*-questions were similar, *wh*-foci's F0 were always higher in *wh*-questions than for *yes/no* questions. Monosyllabic *wh*-foci patterned only with disyllabic *wh*-foci's first syllable in duration and intensity. This suggests that monosyllabic foci were not influenced by the following SFPs. More ChM-TwM differences were found for monosyllabic foci than for disyllabic ones.

Second, for sentential intonation across both groups, statements F0 contours exhibited initial prominence, with a gradual lowering toward the end of the sentence. However, for both types of questions, F0 rose dramatically for *wh*-phrases, as well as for some of SFPs. This unexpected rise in F0 toward the sentence-end indicates that prosodic organization not only takes account of the location of *wh*-phrases, but also the syntactic-semantic information expressed by SFPs. Regional differences were also found. ChM clearly distinguished among three sentence types through F0 contours, while TwM only distinguished statements from questions.

Third, information structure prosodically accommodated for syntactic information. Regarding intensity, our results support the findings of Chen *et al.*[20] about younger speakers of TwM: i.e., that both ChM speakers and younger TwM speakers perform post-focus reduction. However, rather than the on-focus lengthening effects reported in the existing literature, we found that the final syllables of *wh*-focus were sometimes significantly shorter than those found in statements. We also observed unexpected lengthening and higher intensity in question SFPs. Considering the function of SFP—defining sentence types—our findings suggest that the existence of SFPs requires the overall prosodic organization to comply with the representations of sentence types. This seems to be the most likely reason that acoustic variation by sentence type was observed among our tone 1 SFPs. Such a phenomenon could also explain why there was no post-focus reduction of F0 for SFPs.

Our results are also consistent with a claim made by Richards,[27] that, *wh in situ* is allowed, owing to the support from the prosodic phrasing formed by *wh*-interrogatives and their associated complementizer heads (SFPs). While the exact mechanism whereby prosodic phrasing interacts with syntax and word length in speech still calls for further investigation, this study provides important preliminary evidence of the interaction of prosody, information structure, semantics, and syntax in speech production. We also expect that similar interactions will be observable in other tonal languages. Our plans for future research involve examining these effects among learners of tone languages, as well as investigating how Mandarin listeners perceive and use

prosodic differences when interpreting sentences, in order to advance our cross-linguistic understanding of prosodic mechanisms in communication.

Acknowledgments

References and links

[1] S.-Y. Kuroda, "Generative grammatical studies in the Japanese language," Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, MA, 1965.

[2] Y.-h. A. Li, "Indefinite wh in Mandarin Chinese," J. East Asian Ling. **1**(2), 125–155 (1992).

[3] C. N. Li and S. A. Thompson, *Mandarin Chinese: A Functional Reference Grammar* (University of California Press, Berkeley, CA, 1981).

[4] S.-A. Jun and M. Oh, "A prosodic analysis of three types of wh-phrases in Korean," Lang. Speech **39**(1), 37–61 (1996).

[5] M. Deguchi and Y. Kitagawa, "Prosody and wh-questions," in *Proceedings of NELS* (2002), Vol. 1, pp. 73–92.

[6] S. Ishihara, "Intonation and interface conditions," Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, MA, 2003.

[7] H. Truckenbrodt, "An analysis of prosodic f-effects in interrogatives: Prosody, syntax and semantics," Lingua **124**, 131–175 (2013).

[8] J. Yun, "Meaning and prosody of wh-indeterminates in Korean," Ling. Inq. **50**(3), 630–647 (2019).

[9] F. Hu, "A prosodic analysis of wh-words in standard Chinese," in *Speech Prosody 2002*, Aix-en-Provence, France (April 11–13, 2002).

[10] S. Shyu and T.-L. Tung, "Mandarin wh-phrases and prosody," Stud. Pros. Grammar **3**, 32–76 (2018).

[11] L. Xu, "Manifestation of informational focus," Lingua **114**(3), 277–299 (2004).

[12] X.-n. S. Shen, *The Prosody of Mandarin Chinese* (University of California Press, Berkeley, CA, 1990), p. 118.

[13] S. Feng, *Hanyu yunlü jufaxue (zengding ben)* (*Chinese Prosodic Syntax*), Studies in Prosodic Grammar Vol. 3, edited by S. Feng and Q. Ma (Beijing Language and Culture University Press, Beijing, 2013).

[14] S. Jin, "An acoustic study of sentence stress in Mandarin Chinese," Ph.D. thesis, The Ohio State University, Columbus, OH, 1996.

[15] Y. Xu, "Effects of tone and focus on the formation and alignment of f0 contours," J. Phon. **27**(1), 55–105 (1999).

[16] Y. Xu, C. Szu-Wei, and W. Bei, "Prosodic focus with and without post-focus compression (PFC): A typological divide within the same language family?," Ling. Rev. **29**(1), 131–147 (2012).

[17] Y. Chen, Y. Xu, and S. Guion-Anderson, "Prosodic realization of focus in bilingual production of Southern Min and Mandarin," Phonetica **71**, 249–270 (2014).

[18] Y.-Y. Hsu and A. Xu, "Focus acoustics in Mandarin nominal," in *INTERSPEECH* (2017), pp. 3231–3235.

[19] S.-w. Chen, B. Wang, and Y. Xu, "Closely related languages, different ways of realizing focus," in *Tenth Annual Conference of the International Speech Communication Association* (2009).

[20] Y. Chen and C. Gussenhoven, "Emphasis and tonal implementation in standard Chinese," J. Phon. **36**(4), 724–746 (2008).

[21] Y.-Y. Hsu, A. Xu, and H. Ngai, "Focus prosody in Cantonese and Teochew noun phrases," in *Proceedings of 9th International Conference on Speech Prosody* (2018), pp. 961–965.

[22] R. K.-Y. Tsui and S. X. Tong, "An acoustic study on the effect of the interaction between intonation and lexical tones on the realization of Cantonese sentence-final particles," in *Proceedings of the 9th International Conference on Speech Prosody* (2018).

[23] W. L. Wu, "Sentence-final particles in Hong Kong Cantonese: Are they tonal or intonational?," in *Tenth Annual Conference of the International Speech Communication Association* (2009).

[24] Y. Xu, "ProsodyPro–A tool for large-scale systematic prosody analysis," in Proceedings of the TRASP Conference (2013), pp. 7–10.

[25] C. Gu, *Smoothing Spline ANOVA Models* (Springer Science & Business Media, New York, 2013).

[26] D. Bates, M. Mächler, B. Bolker, and S. Walker, "Fitting linear mixed-effects models using lme4," J. Stat. Softw. **67**(1), 1–48 (2015).

[27] N. Richards, *Uttering Trees* (MIT Press, Cambridge, MA, 2010).