Journal of Phonetics 95 (2022) 101190

Contents lists available at ScienceDirect

Journal of Phonetics

journal homepage: www.elsevier.com/locate/Phonetics

Research Article

The prosodic marking of rhetorical questions in Standard Chinese

Katharina Zahner-Ritter [a,b,*], Yiya Chen [c], Nicole Dehé [b], Bettina Braun [b]

[a] University of Trier, Phonetics, Universitätsring 15, 54296 Trier, Germany
[b] University of Konstanz, Department of Linguistics, P.O. Box 186, 78457 Konstanz, Germany
[c] Leiden University Center for Linguistics, Postbus 9515, 2300 RA Leiden, the Netherlands

ARTICLE INFO

Article history:
Received 12 January 2021
Received in revised form 1 September 2022
Accepted 7 September 2022

Keywords:
Lexical tone
Prosody
Intonation
Rhetorical questions
Standard Chinese

ABSTRACT

The present study investigates the prosody of information-seeking (ISQs) and rhetorical questions (RQs) in Standard Chinese, in polar and *wh*-questions. Like in other languages, ISQs and RQs in Standard Chinese can have the same surface structure, allowing for a direct prosodic comparison between illocution types (ISQ vs RQ). Since Standard Chinese has lexical tone, the use of *f0* as a cue to illocution type may be restricted. We investigate the prosodic differences between ISQs and RQs as well as the interplay of prosodic cues to RQs. In terms of *f0*, results showed that RQs were lower in *f0*, with the *f0* range on the first word being expanded followed by *f0* compression. RQs were further longer in duration and more often realized with non-modal voice quality (glottalized voice) as compared to ISQs. These prosodic cues were largely manipulated in tandem (illocutionary pairs with larger durational differences also showed larger differences in mean *f0*; voice quality, in turn, seemed to be an additional cue). We suggest three possible explanations (assertive force, focus, speaker attitude) that unite the present findings on RQs in Standard Chinese with the findings on RQs in other, non-tonal languages.

1. Introduction

Questions such as *Who eats lemons?* may serve more than one function in discourse, two of which are of interest here. First, as an information-seeking question (henceforth ISQ), the interrogative aims at requesting information from the interlocutor. Second, as a rhetorical question (henceforth RQ), it serves to make a point, here to convey that nobody eats lemons (Biezma & Rawlins, 2017; Caponigro & Sprouse, 2007; Han, 2002 on ISQ vs RQ).[1] The present paper investigates the prosodic differences between string-identical ISQs and RQs in Standard Chinese, as well as the interplay between or combination of the cues to RQs. We compared the prosodic realizations in two question types, namely polar questions (1) and constituent (henceforth *wh*-) questions (2). Understanding the prosodic realization of RQs in Standard Chinese will help us gain a broader and generalizable knowledge of question types across languages and their realizations.

(1)	有人(yǒurén)	吃(chī)	柠檬(níng méng)	吗 (*ma*)/么(*me*)?
	Anyone	eat	lemon	sentence-final particle
	'Does anyone eat lemons?'			
(2)	谁(shéi)	吃(chī)	柠檬(níng méng)?	
	Who	eat	lemon	
	'Who eats lemons?'			

As (1) shows, in Standard Chinese, polar questions typically use particles such as 吗*ma*, 么*me*, and 吧 *ba* that mostly occur in sentence-final position (Chao, 1968; Liing, 2014). The sentence-final particle 么*me*,[2] for instance, can turn a statement

* Corresponding author at: University of Trier, Phonetics, Universitätsring 15, 54296 Trier, Germany.
E-mail addresses: k.zahner-ritter@uni-trier.de (K. Zahner-Ritter), yiya.chen@hum.leidenuniv.nl (Y. Chen), nicole.dehe@uni-konstanz.de (N. Dehé), bettina.braun@uni-konstanz.de (B. Braun).

[1] Rhetorical questions do not always imply an empty set as answers (here: 'nobody likes lemons'), but their answers may also imply a specific entity that is known or inferable to the interlocutor (see e.g., the example in Biezma & Rawlins, 2017, p. 304, in which the rhetorical questions implies the answer 'Luca': 'You should stop saying that Luca didn't lie the party last night. After all, who was the only one that was still dancing at 3am?'). The present paper focuses on the former type of rhetorical questions in which the implied answer is the empty set.

[2] Note that speakers vary in the exact pronunciation of the particle 么, particularly with respect to the quality of the vowel that follows the nasal [m], varying between *me* and *ma*. In our experiment, participants produced the sentence-final particle 么 as *ma* [ma] or *me* [mə] about equally often (49.5% *ma*, 50.5% *me*). There was no difference in the distribution of vowel quality between illocution types (p_{adj} = 0.23).

https://doi.org/10.1016/j.wocn.2022.101190
0095-4470/

into a polar question. 么 *me* may be used in both ISQs and RQs. Regarding *wh*-questions, Standard Chinese is a *wh*-in-situ language (Cheng, 1991). That is, *wh*-elements surface in their syntactic base position. In our study, we only use *wh*-questions in which the *wh*-word is the subject and thus precedes verb and object. For both polar and *wh*-questions, the same string of words can be used to produce a question with an information-seeking or a rhetorical illocution. Clearly, polar questions differ from *wh*-questions with respect to their semantics and syntax (Groenendijk & Stokhof, 1984, pp. 1744–1747; Krifka, 2011), cf. (1) and (2). Also, in our experiment, the sentence-final particle 么 *me* occurs in polar questions only, leading to differences in the number of syllables across *question type*. Crucially, the main comparison of our study is between ISQs and RQs, which essentially are string-identical within each question type (polar vs *wh*-questions).

Previous research on a variety of different languages has shown that ISQs and RQs differ in a number of prosodic characteristics: *f0*, duration, and voice quality (Dehé et al., 2022 for a recent overview). These differences hold true for both polar and *wh*-questions. In regard to the use of *f0* for the marking of illocution type, lexical tone languages provide a particularly interesting test case for the distinction between ISQs and RQs, since tone is contrastive at the lexical level and phrase-level *f0* is therefore constrained by the canonical form of the lexical tone (e.g., Gussenhoven, 2004). Studying the marking of illocution type in tone languages hence allows us to contribute to our understanding of the use of *f0* beyond its primary function of marking lexical tone (cf. Chang, 1975). In the present paper, we focus on Standard Chinese by speakers who were born and grew up in Beijing with Standard Chinese – a variety in which every syllable carries one of four lexical tones, or the neutral tone (Chao, 1930, 1956; Chen, 2016, 2022; Lin, 2007), see below in Section 1.2.

In the remainder of Section 1, we first summarize the main findings on the prosodic differences of ISQs vs RQs for the languages on which experimental studies have been conducted (Section 1.1). Section 1.2 reviews the literature on the use of *f0* and other prosodic cues for lexical and non-lexical purposes in Standard Chinese. Based on this background, Section 1.3 outlines the research questions and hypotheses. Section 2 presents the methods of the production experiment, Section 3 its results. Section 4 discusses the results in the context of acoustic cue weighting and other non-lexical uses of prosody in Standard Chinese. From a broader, cross-linguistic context, it also includes findings from typologically different languages to discuss cross-linguistic signals of rhetorical questions (Section 4.1). We finally discuss implications of the interplay between prosodic cues to RQs for the modelling of the relation between prosody and meaning (Section 4.2), and conclude in Section 5.

1.1. Rhetorical questions and how they are signalled

ISQs constitute a directive speech act of requesting information from the addressee (Groenendijk & Stokhof, 1984; Krifka, 2011), also known as neutral, real, or genuine questions. RQs may share their syntactic surface form with ISQs but attempt to commit the interlocutor to the answer that is presupposed in the RQ (cf. Biezma & Rawlins, 2017; Han, 2002, p. 202, who considers RQs as assertions), e.g., *nobody likes lemons* in examples (1) and (2). Signals to rhetorical illocution are, among others, shared world-knowledge, e.g., *Is the Pope catholic?* (Han, 2002, p. 216), syntactic cues such as strong negative polarity items, e.g., “ever” in *What has John ever done for Sam?* (Han, 2002, p. 202), or lexical cues such as discourse particles, e.g., German *schon* (Bayer & Obenauer, 2011, p. 454; see also Dehé, Wochner, & Einfeldt, 2022). For Standard Chinese, B. Xu (2013) argues that questions containing 难道 *nándào* necessarily have a rhetorical illocution, see (3). Fang (2021) has recently also argued that 呢 *ne* may signal a contradiction between the stated message and an existing assumption, giving rise to a rhetorical illocution.

(3)	难道	谁	帮过	你	吗?
	Nándào	shuí	bāng-guò	nǐ	ma?
	Nandao	who	help-EXP	you	PRT

‘Who helped you?’ (=No one helped you.), cf. Xu (2013, p. 509)

Given that these cues are optional, an interrogative may be ambiguous between ISQ or RQ meaning in Standard Chinese. Previous work has shown that prosodic cues can distinguish string-identical ISQs and RQs in production, particularly in regard to three prosodic dimensions, i.e., *f0*, duration, and voice quality (Dehé et al., 2022, for overview), but most of these studies focused on intonation languages, for which the following has been found:

- RQs are longer (or realized with a slower speaking rate) as compared to ISQs in a variety of different languages – including German (Braun et al., 2019; Braun, Einfeldt, Esposito, & Dehé, 2020), English (Dehé & Braun, 2020b), Icelandic (Dehé, Braun, & Wochner, 2018; Dehé & Wochner, 2022), French (Beyssade & Delais-Roussarie, 2022), Italian (Sorianello, 2018, 2019), and Estonian (Sahkai, Asu, & Lippus, 2022).
- For German, English, and Icelandic, more instances of breathy voice have been found for RQs as compared to ISQs (Braun et al., 2019; Dehé & Braun, 2020b; Dehé & Wochner, 2022), while for Estonian, more instances of glottalized voice have been observed for RQs compared to ISQs (Sahkai et al., 2022).
- RQs also differ from ISQs in the position of the nuclear pitch accent (Dehé & Braun, 2020b for English; Sahkai et al., 2022 for Estonian), the type of pitch accent (Beyssade & Delais-Roussarie, 2022 for French; Braun et al., 2019 for German; Dehé & Braun, 2020a for Icelandic; 2020b for English; Sorianello, 2019 for Italian), and the types of final edge tones (Beyssade & Delais-Roussarie, 2022 for French; Braun et al., 2019 for German; Dehé & Braun, 2020b for English; Sahkai et al., 2022 for Estonian; Sorianello, 2019 for Italian).

In Japanese, a pitch accent language, RQs have been shown to be longer and lower in overall *f0* than ISQs; initial lowering is furthermore a strong perceptual indicator for RQs in Japanese (Miura & Hara, 1995).

Research on the prosody of RQs has only recently included tone languages – but has so far been restricted to one question type: *wh*-questions (Lo & Kiss, 2020; Lo, Kiss, & Tulling, 2019b). Tone languages are particularly interesting since they pose questions for the interaction between lexical tone and post-lexical intonation in the marking of illocution type (Chen, 2022 for overview). For Cantonese, a tone language with six lexical tones (Zhang, Duanmu, & Chen, 2021 for overview),

Lo, Kiss, and Tulling (2019a) investigated the sentence-final particles in *wh*-questions and found them to be longer and lower in RQs than in ISQs. Lo and Kiss (2020) furthermore studied *wh*-questions in Mandarin Chinese and found *wh*-RQs to be overall longer than string-identical ISQs, except for the sentence-final particle. Moreover, the sentence-final particle was lower in *f0* in RQs as compared to ISQs, and more often realized with glottalized voice; in turn, the *wh*-word was higher in RQs as compared to ISQs. No *f0* differences have been reported for the middle part of the sentence.

Hence, lexical tone languages distinguish different illocution types in *wh*-questions using duration, *f0*, and – for sentence-final particles – voice quality. So far, nothing is known on whether this prosodic marking generalizes to polar questions, whether it is limited to the realization of the *wh*-word and the sentence-final particle, and whether speakers use the different kinds of cues to RQs in a compensatory manner (trading relation) or in tandem (cf. Schertz & Clare, 2019). For tone languages, in particular, it might be the case that adjustments in *f0* (given its primary function of marking lexical tone) are limited and get compensated by larger differences in other prosodic cues. Such limited adjustments in *f0* have been observed by Chen and Gussenhoven (2008) for the realization of different levels of emphasis. In particular, when speakers of Standard Chinese were encouraged to produce different levels of emphasis, they tended to lengthen more without further modification of the *f0* range. In the present study, we test this possibility by examining the interplay between different prosodic cues to RQs. As we will briefly summarize in the following section, beyond its lexical function, *f0* also serves post-lexical functions in Standard Chinese (Xu, 2019; Zhang et al., 2021, for overviews).

1.2. Lexical and non-lexical functions of f0 and other prosodic cues in Standard Chinese

Standard Chinese is a tone language in which every syllable carries one of four lexical tones: Tone 1 (T55, high-level, ē), Tone 2 (T35, rising, é), Tone 3 (T214, low-rising, ě) and Tone 4 (T51, falling è), or the neutral tone (Chao, 1930, 1956; Chen, 2016, 2022; Lin, 2007).[3] Tone 3 is also frequently associated with creaky voice (e.g., Chao, 1956, p. 53; Kuang, 2017, and references therein, p. 1694). *F0* hence primarily serves a lexical function in Standard Chinese, such that a change in lexical tone leads to a change in lexical meaning. The canonical shape of the tones is directly evident when tones are produced in isolation, and tones in multi-syllabic phrases are typically influenced by preceding or following tones (e.g., Shen, 1990; Xu, 1997; Xu & Liu, 2006). In addition to the lexical function, *f0* (and other prosodic cues) are also used to convey post-lexical functions (Xu, 2019; Zhang et al., 2021, for overviews).

At the post-lexical level, *f0* is used to mark information structure, speech acts or affective states (Chen, 2022; Xu, 2019; Zhang et al., 2021, for overviews). Regarding information structure, tones on focused words are typically realized with a greater *f0* range than words in non-focal position (Jin, 1996; Liu & Xu, 2005; Xu, 1999; Chen and Braun, 2006) and with longer durations (Chen, 2006; Chen & Gussenhoven, 2008; Jin, 1996; Xu, 1999). Tonal and segmental contrasts are hyperarticulated with respect to their distinctive characteristics (Chen, 2008; Chen & Gussenhoven, 2008). Generally speaking, focus-induced *f0* adjustments have been reported to apply to the whole utterance, with the focused element being expanded in *f0* and the region thereafter being compressed,[4] a mechanism termed post-focal compression (Gårding, Zhang, & Svantesson, 1983; Jin, 1996; Xu, 1999; Xu & Xu, 2005; Zhang et al., 2021, for overview).

Beyond focus, prosodic cues are used to convey emotions or affective states (Li, Fang, & Dang, 2011; Liu & Pell, 2012; Yuan, Shen, & Chen, 2002). “Disgust”, for instance, is associated with a lowering in *f0* (Li et al., 2011; Liu & Pell, 2012), a slower speech rate (Liu & Pell, 2012), and low harmonics-to-noise ratio values (Liu & Pell, 2012), hinting to the use of non-modal voice quality for this emotion (Keating, Garellek, & Kreiman, 2015).

Importantly, prosody, in particular *f0* and duration, also marks speech acts, such as the difference between statements and questions in string-identical utterances, see (4).

(4)	有人(yǒurén)	吃(chī)	柠檬(níngméng)
	Anyone/Somebody	eat	lemon
	‘Somebody eats lemons.’ or ‘Does anyone eat lemons?’		

Specifically, polar-ISQs are globally produced with higher overall *f0* than string-identical declaratives (Lee, 2005; Liu & Xu, 2005; Yuan, 2006), with the difference in *f0* becoming larger towards the end of the utterance (Yuan, 2006). Contrary to Cantonese, where questions end in a final rise irrespective of the lexical tone, Standard Chinese is more faithful to the shape of the lexical contour at the end of the utterance. This means, for instance, that the falling Tone 4 is still a falling tone in questions, but the range of the fall is reduced; conversely, Tone 2 is rising also in questions, but compared to declaratives, it is realized with an enhanced *f0* range (Chen, 2022; Zhang et al., 2021). In terms of duration, except for the last syllable, syllables have been shown to be shorter in polar questions than in declaratives (Yuan, 2006).

Prosody also distinguishes between *wh*-questions and string-identical declaratives. In *wh*-questions, *shénme* is the *wh*-pronoun ‘what’, while in declaratives, together with the licensor *diānr* (‘a little’), it is an indefinite / existential, meaning ‘a little bit of something’ (cf. Yang, 2018). In production, *f0* and other prosodic cues distinguish between the two readings: For example, in sentences containing *shénme*, *wh*-ISQs exhibit higher *f0* compared to string-identical declaratives, mostly towards the end of the utterance (Liu, Li, & Jia, 2016; Yang, 2018; Yang, Gryllia, & Cheng, 2020). Yang (2018) further shows an increased *f0* range in *shénme* ‘what’ for questions. Additionally, utterance and word durations are shorter in *wh*-ISQs than in declaratives, with an exception of *shénme* 'what',

[3] The numbers (e.g., 51 for the falling Tone 4) indicate the pitch levels involved in the tonal movement, with 1 being at the low end and 5 at the high end; diacritics (e.g., ē) are placed on vowel (here ‘e’) and indicate the direction of the tonal movement, see International Phonetic Alphabet for all diacritics, see https://www.internationalphoneticassociation.org/content/full-ipa-chart (last access: 18 November 2021).

[4] The *f0* expansion and compression effects of focus are also constrained by the tonal context (e.g., Chen, 2010) and prosodic structure of the focused element (see review in Chen, Lee, & Pan, 2016).

for which the pattern was reversed (Yang, 2018); see also Yang et al. (2020).[5]

Taken together, Standard Chinese uses *f0* modifications to express post-lexical functions on top of lexical tone – a phenomenon that has been referred to as the "multiplexing of the f0 channel (Zhang et al., 2021, p. 9, chapter 24.5). In addition to *f0*, other prosodic parameters such as duration and voice quality also serve post-lexical functions in Standard Chinese, particularly with respect to the marking of focus, sentence type and attitudes. One of the remaining questions is how Standard Chinese employs *f0* and other prosodic cues to differentiate between string-identical ISQs and RQs.

1.3. Research questions and hypotheses

The present study investigates the prosodic marking of RQs as compared to ISQs in Standard Chinese, using prompted productions of target questions (polar and *wh*-questions). We used an experimental paradigm that has been employed for other languages (Braun et al., 2019; Dehé & Braun, 2020a, 2020b) and adapted it to Standard Chinese: Chinese participants read short contexts (which described different situations) followed by target interrogatives. The contexts were created such that they either triggered an ISQ or an RQ illocution. Target questions were string-identical in both readings (ISQ and RQ). Our two main Research Questions (Qs) are the following:

- **Q1:** Do string-identical ISQs and RQs in Standard Chinese prosodically differ from each other, both in polar and in *wh*-questions, and if so, what are the prosodic cues?
- **Q2:** Are prosodic cues that distinguish illocution type used in a compensatory manner or are they modified in tandem?

With respect to Q1, we analysed *f0*, duration, and voice quality, which appear to be the main cues cross-linguistically (Dehé et al., 2022). Based on previous work (see 1.1), we put forward the following hypotheses:

- **H1a:** RQs will be overall lower in their *f0* trajectory than ISQs.
- **H1b:** RQs will be longer in their duration than ISQs, for all words.
- **H1c:** RQs will more often be realized with non-modal voice quality than ISQs.

With respect to **Q2**, we test whether cues are used in a compensatory manner (trading relation), such that smaller adjustments in *f0* (which is the primary cue to signal tone) are compensated by larger differences in other prosodic cues, such as duration or voice quality (cf. Schertz & Clare, 2019). Here, we focus on illocutionary pairs consisting of an ISQ and its corresponding RQ. A trading relation is present between *f0* and other prosodic cues when *f0* does not differ between an ISQs and its corresponding RQs, but duration and voice quality do. A modification in tandem is present when all cues (*f0*, duration, and voice quality) differ between an ISQ and its corresponding RQ.

[5] The authors in Yang et al. (2020) take the reversal of results for the *wh*-word (i.e., longer duration in *wh*-question reading than in declarative reading) to reflect a difference in focus marking, with *shénme* being focused in the question reading (meaning 'what') while it is not in the declarative reading (meaning 'something').

2. Methods

The experiment was run in Beijing in spring 2018. The study was approved by the Ethics Committee of the University of Konstanz (Institutional Review Board number: IRB 30/2016).

2.1. Materials

Twenty-two polar and 22 *wh*-questions were constructed, along with two contexts for each question (one context eliciting an ISQ reading, the other an RQ reading), resulting in 22 context-question quadruplets, see Table 1. The quadruplets were translated from Braun et al. (2019) by a native speaker of Standard Chinese. To compare the results across languages, we aimed at maximal comparability to our work on the prosody of RQs in German, English, and Icelandic (Braun et al., 2019; Dehé & Braun, 2020a, 2020b) – in terms of a) the semantics of the situation (context) and the structure and content of the target interrogatives, and b) in terms of the statistical power (number of items). We made six changes for cultural reasons and took care to balance lexical tone in the final syllable of the noun, which is the final syllable in the sentence in *wh*-questions and the penultimate in polar questions, which included a sentence-final particle. Polar questions started with *yǒurén* ('anyone') and *wh*-questions with *shéi* ('who'), both followed by a verb, and an object consisting of one noun, plus the particle in polar questions, see Table 1 for an example quadruplet and Appendix A for a complete list of target questions of the present experiment. The verbs and objects in the target questions were of different length (verbs were either mono- or bisyllabic; object nouns consisted of one to five syllables, see Appendix A). The ISQ version of the context always contained the sequence 'you would like to know', in accordance with the property of ISQs that they seek information. By contrast, the RQ version of the context always contained 'it is known that', indicating that the answer is obvious to the interlocutors. Interrogatives were felicitous in both illocution types and contained an object noun that was non-constraining as to one of the readings (e.g., *lemons*), as verified by a pre-test conducted online. In particular, native speakers of Standard Chinese indicated for all 22 items whether they agreed or disagreed with the proposition in the interrogative. On average, participants agreed in 55.3% of the cases (with a range of 13.3% to 83.3% across individual items). Hence, the propositions in the interrogative sentences were on average ambiguous as to one of the two illocution types.

Polar questions always started with the pronominal subject *yǒurén* 'anyone'; *wh*-questions started with the question word *shéi* 'who'. Consequently, within question types, the first word (the subject of the sentence) was always the same – and hence always carried the same lexical tones. Verbs and object nouns varied in terms of number of syllables and lexical tone, see Appendix A. The end of the utterance was balanced for lexical tone: For the object noun, lexical tone was distributed such that all four tones occurred in the last syllable of the noun (six times Tone 1, six times Tone 2, four times Tone 3, and six times Tone 4, in both polar and *wh*-questions). Note that different tones were used for reasons of generalizability and to avoid confounding with respect to the use of voice quality

Table 1
Example context and question quadruplet in an ISQ (left) and RQ reading (right), for a polar (upper panel) and a *wh*-question (lower panel).

Polar question			
Context for ISQ			
At a party, you offer cake made with lemons. You would like to know which of the guests like this fruit and whether they would like some or not.			
You say to your guests:			
有人 (yǒurén)	吃 (chī)	柠檬 (níngméng)	么 (me)?
Anyone	eat	lemon	sentence-final particle
'Does anyone eat lemons?'			
Context for RQ			
Your aunt offers lemons to her guests. However, it is known that this fruit is too sour to be eaten on its own.			
You say to your cousin:			
有人 (yǒurén)	吃 (chī)	柠檬 (níngméng)	么 (me)?
Anyone	eat	lemon	sentence-final particle
'Does anyone eat lemons?'			
Wh-question			
Context for ISQ			
At a party, you offer cake made with lemons. You would like to know which of the guests like this fruit and would like some of it.			
You say to your guests:			
谁 (shéi)	吃(chī)	柠檬(*níngméng)*?	
Who	eat	lemon	
'Who eats lemons?'			
Context for RQ			
Your aunt offers lemons to her guests. However, it is known that this fruit is too sour to be eaten on its own.			
You say to your cousin:			
谁 (shéi)	吃(chī)	柠檬(*níngméng)*?	
Who	eat	lemon	
'Who eats lemons?'			

and duration (Kuang, 2017; 2018, among others). However, due to semantic constraints we only controlled tone at the beginning and end of the utterance. Furthermore, to have comparable conditions, polar and *wh*-questions contained the same predications (e.g., *eating lemons*). The particle 么 *me*, which is commonly used in polar questions (Liing, 2014), was included in both polar-ISQs and polar-RQs. There was no question particle for *wh*-questions. Since we focus on the difference between ISQs and RQs, the difference in structure in polar and *wh*-question is secondary.

Additionally, 34 fillers and their contexts (declaratives with attachment ambiguities, exclamatives, alternative questions and neutral polar questions) were translated from Braun et al. (2019) and used in the present experiment.

2.2. Procedure

Two experimental lists were constructed, each containing both question types (polar and *wh*) and both illocution types (ISQ and RQ). Each list contained half of the polar questions (N = 22; 11 in an ISQ and 11 in an RQ reading) and half of the *wh*-questions (N = 22; 11 in an ISQ and 11 in an RQ reading) and all 34 fillers. Illocution type (ISQ vs RQ) was manipulated *within-subjects*, i.e., each participant produced both the ISQ version and the RQ version of each target interrogative. Each list contained only one question type of each illocutionary pair (ISQ and RQ), either the polar or the *wh*-question. One of the two lists was randomly assigned to each participant. Each participant received a randomized order of items, with the constraint of separating the same question (in the two readings) by at least four other items. Three practice trials preceded 78 trials (44 experimental and 34 fillers). Participants received oral instructions in Standard Chinese by the experimenter, a research assistant who is a native speaker of Standard Chinese. The experiment was controlled in *Presentation* (Presentation, 2000).

On each trial, participants silently read a context displayed on a computer screen. Upon button press, the target interrogative appeared on the screen and the recording started. Participants were instructed to read each context carefully and to produce the subsequent interrogatives in a way that was suitable in the given context. They were allowed to produce the sentence again, if needed. Upon another button press, a new trial started. Productions were recorded using a headset microphone (Shure SM10A) and digitized onto a computer (44.1 kHz, 16 Bit). The experimenter did not interfere during the experiment. Testing took place in a quiet room and the experiment lasted about 25–30 minutes.

2.3. Participants

Ten native speakers of Standard Chinese (all female, average age = 26.5 years; SD = 2.0 years) born and raised in Beijing with Standard Chinese participated in the study. Two additional speakers born and raised elsewhere were excluded from the present analysis in order to minimize potential influence of dialectal variation.

2.4. Data preparation and annotation

In total, 440 target interrogatives were produced (44 contexts × 10 participants). Twenty-two interrogatives (3.9%) were excluded from the analysis because of technical errors (N = 2), mispronunciations (N = 8), or pauses / hesitations between the words (N = 12). The final data set (N = 418) comprised 212 polar questions (106 ISQs, 106 RQs) and 206 *wh*-questions (103 ISQs, 103 RQs).

All interrogatives were annotated in Praat (Boersma and Weenink, 2016) on three tiers: on the syllabic level, the word level, and with respect to voice quality (at the beginning and end of the utterance), see Fig. 1. Segmental boundaries were manually placed by a native speaker of Standard Chinese based on standard segmentation criteria (Turk, Nakai, & Sugahara, 2006). Pitch tracking errors were manually corrected (first author) by removing erroneous pitch points in the Praat Manipulation editor (Boersma & Weenink, 2016) and saving the modified Manipulation-Object as a wav-file (Pitch overlap-add). The corrected files were used for further processing and analyses. *F0* values of the *f0* trajectory over time were automatically extracted from the files with corrected *f0* using the Praat script ProsodyPro (Y. Xu, 2013). Specifically, we extracted ten measurements (in Hz) from each word.[6]

[6] Note that ProsodyPro uses interpolation for unvoiced portions of the signal. From our perspective, this is entirely unproblematic for our analyses, given that the voiceless sounds in our materials are mostly voiceless fricatives, which have been show to carry *f0* information for the human perceptual system ("segmental" intonation, Niebuhr, 2012, 2017). Importantly, voiceless portions are the same in both illocution types (ISQ and corresponding RQ), which is the main interest of our study.

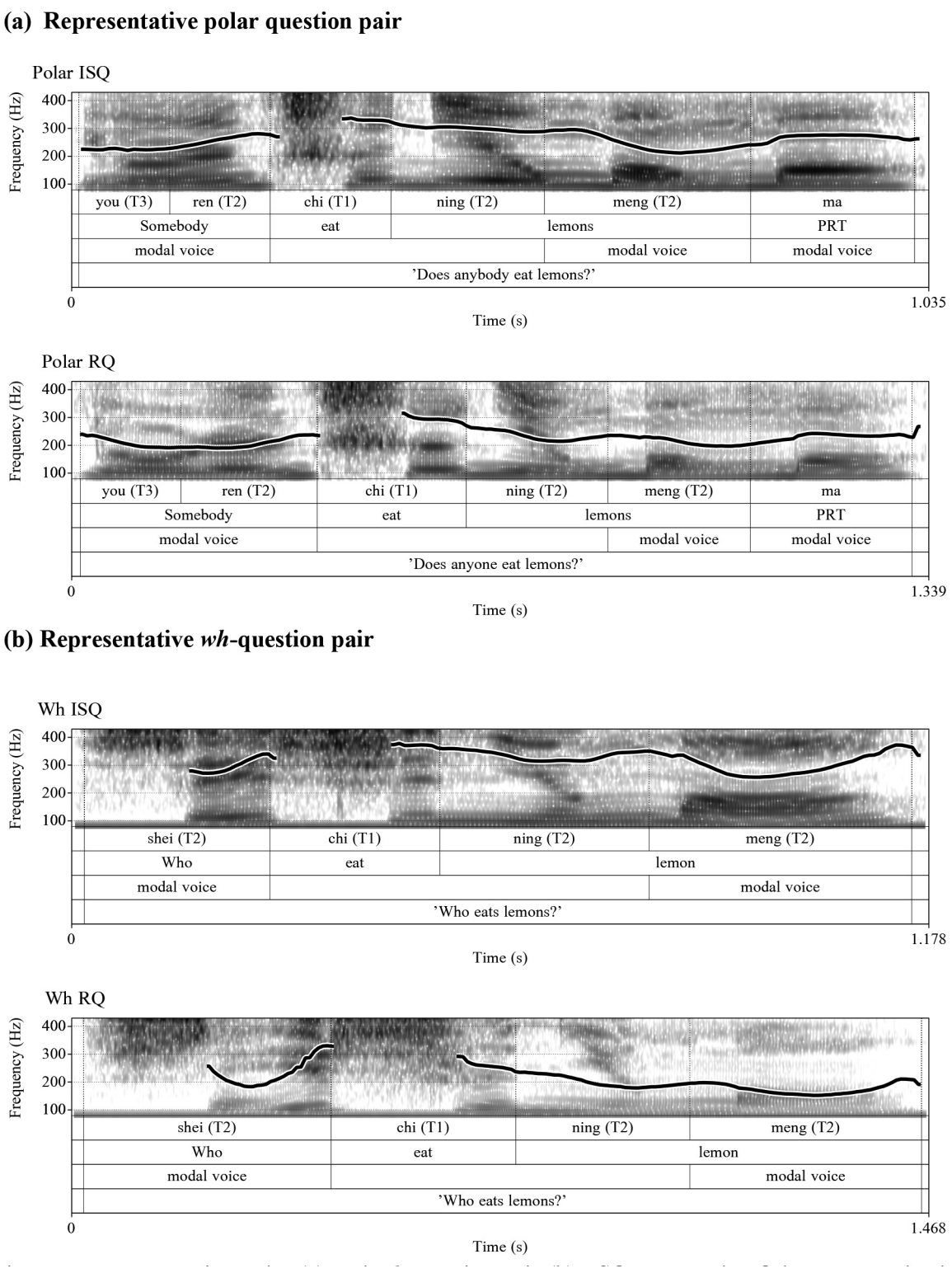

Fig. 1. Representative polar (a) and *wh*-question pair (b), ISQ top panel, RQ bottom panel. Tier 1 shows the syllable representation in Standard Chinese with the respective information on lexical tone; tier 2 gives the English word-by-word translation; tier 3 indicates the voice quality classification at the first word and the last syllable(s); tier 4 provides the English translation of the question.

Voice quality was annotated by two separate annotators (first author and native Chinese student assistant), based on perceptual classification, as modal, breathy, or glottalized voice quality (Braun et al., 2019; Laver, 1980). To avoid effects of tonal identity on voice quality (Kuang, 2017; 2018, among others), we labelled voice quality only at the beginning and the end of the utterance, where we controlled or balanced for lexical tone:

- Polar questions: one label for *yǒurén* 'anyone',[7] one label for the last syllable of the object noun, and one for the sentence-final particle (i.e., three voice quality labels in total).
- *Wh*-questions: one label for *shéi* 'who' and one for the last syllable of the object noun (i.e., two voice quality labels in total).

In total, 1048 voice quality labels were set (N = 636 in polar question, i.e., three labels in 212 analysed productions, and N = 412 in *wh*-questions, i.e., two labels in 206 analysed productions). We checked the reliability of the voice quality labelling based on 39.9% of the voice quality labels (i.e., 418 labels). The two labellers agreed in 98.1% of the cases, κ = 0.95, "almost perfect" (Cohen, 1960; Gamer, Lemon, Fellows, & Singh, 2012; Landis & Koch, 1977).

2.5. Statistical analysis

This section gives an overview of the statistical analyses, which were done separately for each question type (polar and *wh*-questions). We also plot the results separately for each question type, even though the effects of illocution type were independent of question type, i.e., are comparable in both question types. Analysis scripts are available at Mendeley upon publications (https://data.mendeley.com/datasets/49nvs73y35/2). Statistical analyses were done in R (R Development Core Team, 2015).

Research Question 1 (prosodic differences between ISQs and RQs, separately for each question type): The main analysis concerned the global prosodic differences between illocution types (ISQ vs RQ) in polar and *wh*-questions with respect to the three prosodic parameters: a) the *f0* trajectory over the target question, b) utterance (and word) durations, and c) voice quality at the beginning and end of the target question (cf. Q1, H1a-c).

To investigate differences in the *f0* trajectory over the target question between ISQs and RQs (cf. H1a), general additive mixed modelling was applied (GAMMs, Wieling, 2018; Wood, 2006, 2017), separately for question type (polar and *wh*-questions). We extracted ten *f0* values for each word in a question, and the resulting time-normalized *f0* contours were compared across *illocution type* (ISQ vs RQ). GAMMs were chosen for the analysis of the *f0* trajectory as they represent an optimal way for the analysis of time-varying data with non-linear relationships and auto-correlation (Baayen, van Rij, de Cat, & Wood, 2018; Wieling, 2018; for a comparison of intonation contrats using GAMM, see Zahner-Ritter, Einfeldt, et al., 2022; Zahner-Ritter, Zhao, Einfeldt, & Braun, 2022). In brief, GAMMs model non-linear dependencies in *f0* and *illocution type* over time via smooth functions. These smooth functions include a pre-specified number of base functions of different shapes, e.g., linear and parabolic functions of different complexity (e.g., Wieling, 2018). Fixed effects are modelled in the same way as in linear mixed effect regression models. In addition, GAMMs also model non-linear effects over time. The visualization of the predicted differences gives the time period in which two contours differ as a function of *illocution type*. For model fitting of the GAMMs, we used the R package *mgcv* (Wood, 2011, 2017); the package *itsadug* was used to plot the model results (van Rij, Wieling, Baayen, & van Rijn, 2017). The response variable was the *f0* value (in Hz) at different time points (10 measurements per word). One model was fitted for polar, one for *wh*-questions. The models included *illocution type* as a parametric effect (fixed effect), along with a factor smooth for the interaction of *illocution type* over (normalized) time, s(Normtime, by = *illocution type*), using the thin-plate regression spline ('tp'). We modelled separate smooths for subjects and items to account for the experimental structure. The model including the smooth term that captured the interaction of *illocution type* over time was subsequently compared to a simpler model without the smooth term, using the function CompareML(). This comparison tested whether the inclusion of this term significantly improved the fit of the model in terms of Maximum Likelihood (see Porretta, Tucker, & Järvikivi, 2016). All models were corrected for autocorrelation in the data using a correlation parameter, determined by the acf_resid()function from the package *itsadug* (van Rij et al., 2017). Model fits were checked using the gam.check() function and the number of base functions (k) was adjusted if necessary. Also, best-fitting models were re-run with the scaled *t* distribution (family = "scat"), closely following the suggestion in van Rij et al. (2019, p. 17), in order to account for tailed residuals.

Utterance (and word) durations (cf. H1b) were statistically analysed using linear mixed-effects regression models (lmers, Baayen, Davidson, & Bates, 2008). voice quality labels (cf. H1c) were analysed using logistic mixed effects regression models (*glmers*), coding glottalized voice as 1 and modal voice as 0 (breathy voice did not occur in the data). For *glmers*, we used the "bobyqa"-optimizer (Powell, 2009) in the glmerControl function in order to reduce convergence issues. Otherwise, the modelling procedure was the same for the continuous (duration, H1b) and categorical data (voice quality, H1c): Levels in all categorical variables were dummy coded, i.e., each individual level is compared against the reference level (intercept). *Participants* and *items* were entered as crossed random factors (Baayen et al., 2008). Random slopes were added and retained if they improved the fit of the model (Matuschek et al., 2017) – based on model comparisons with the *anova()* function that compares LogLikelihoods. P-values were obtained using the Satterthwaite approximation implemented in the R package *lmerTest* (Kuznetsova, Brockhoff, & Christensen, 2017). They were adjusted based on the Benjamini-Hochberg correction (Benjamini & Hochberg, 1995) to counteract an increase in type-I-error rate. We report adjusted p-values in the results part (p_{adj}) and assume a standard significance level of 0.05.

Research Question 2 (interplay between cues to RQs): To test whether mean *f0* and utterance duration compensate for each other, we correlated the difference in mean *f0* (Δf0) with

[7] For the disyllabic pronominal subject *yǒurén*, we assigned the respective label ("glottal" or "breathy") if one of the two syllables was non-modal; otherwise, the label "modal" was assigned.

the difference in interrogative duration (Δduration) for each illocutionary pair (ISQ and corresponding RQ by the same speaker). A negative correlation was taken to indicate compensation between cues, i.e., the larger the difference across illocution types in one cue, the smaller the difference in the other. A positive correlation was taken to suggest that the cues are modified in tandem, i.e., the larger the difference in one cue, the larger the difference also for the other cue. Each illocutionary pair was further coded with respect to whether or not it was marked for voice quality. We assigned 'yes' if and only if one of the positions labelled for voice quality (first word, last syllable in object noun, plus sentence-final particle in polar questions) had modal voice in the ISQ version and glottalized voice in the corresponding RQ *and* none of the positions in the respective illocutionary pair showed the reversed contrast (i.e., glottalized in ISQ and modal in RQ). In all other cases, we assigned 'no'. We subsequently checked the distribution of voice quality marking ('Yes') across illocutionary pairs. If voice quality marking compensates other cues (*f0* and duration), we expect voice quality distinctions to occur for illocutionary pairs with weak marking of *f0* and duration (small differences across illocution types).

3. Results

In this section, we will first provide the analyses in response to the first research question (prosodic differences between ISQs and RQs in polar and *wh*-questions), presenting the acoustic comparison for the *f0* trajectory of the target interrogatives (H1a, Section 3.1.1), (word) durations (H1b, Section 3.1.2), and the use of voice quality (H1c, Section 3.1.3). We will then turn to the interplay between these cues to RQs in response to the second research question (Section 3.2).

3.1. Prosodic differences between ISQs and RQs in polar and wh-questions: F0, duration, and voice quality

3.1.1. Global f0 trajectory

Fig. 2 provides a visualization of the time-normalized *f0* trajectory for the entire target interrogatives in the two illocution types (ISQ vs RQ) to reveal global trends in *f0*, for polar (A) and *wh*-questions (B).

Note that Fig. 2 averages the *f0* contours over each word, irrespective of lengths and tones of this word. However, since we are comparing sentence *pairs* (i.e., the same sentences by the same speakers in the two illocution types), we can nevertheless analyse and interpret these average contours. To statistically corroborate the differences in the *f0* trajectory between ISQs and RQs, we used GAMMs. The factor smooths for the interaction of *illocution type* over (normalized) time were necessary for both polar questions ($\chi^2(2.00) = 201.2$, $p_{adj} < 0.0001$) and *wh*-questions ($\chi^2(2.00) = 52.6$, $p_{adj} < 0.0001$), indicating that the impact of illocution type differed over the course of the utterance.

For both polar and *wh*-questions, the final GAMM included *illocution type* as a parametric effect (fixed effect), along with a factor smooth for the interaction of *illocution type* over (normalized) time, s(Normtime, by = *illocution type*) and a smooth for subjects and items (random slopes). The final model accounted for 69.6% of the deviance in polar questions and 65.4% in *wh*-questions. The final model was corrected for auto-correlation as well as re-run with the scat-linking function.[8] Given that we can interpret the GAMM results more intuitively with visualizations, we present the visualized model output in Fig. 3. The summary table of the final model can be found in the supplementary analysis script. Fig. 3 shows the predicted *difference* in *f0* (predicted *f0* values in RQ condition minus ISQ condition). The left panel shows the predicted difference curve for polar questions; the right panel for *wh*-questions. The predicted difference curves show *when* in time ISQs and RQs differ significantly from each other (values below 0 indicate that RQs are lower, values above 0 the reverse; for the period(s) when the 95% confidence interval of the difference curve does not include the horizonal line at zero, the difference is significant, as indicated by the vertical red lines that highlight significant periods).

The *f0* difference curves in Fig. 3 reveal that RQs are lower for most parts of the question (except for parts of the verb). The analyses further reveal a larger *f0* range for RQs on the first word both for *yǒurén* 'anyone' in polar questions and *shéi* 'who' in *wh*-questions. The increased *f0* range seems to be due to a lowering of the low tonal target (around Normtime 5 for polar and Normtime 7 for *wh*-questions, see Fig. 2), resulting in a dip of the contour for RQs. There is no difference across illocution type for almost half of the verb in polar questions (around Normtime 10–15) and for almost one third of the verb in *wh*-questions (around Normtime 11–13). From the verb onwards, the *f0* trajectory between ISQs and RQs diverges more and more as the interrogative unfolds.

To summarize the findings regarding *f0*, Standard Chinese RQs show a lower *f0* trajectory than ISQs, which holds for both polar and *wh*-questions. The *f0* trajectory in RQs is characterized by a larger *f0* range for the subject word (*yǒurén* 'anyone' in polar and *shéi* 'who' in *wh*-questions), due to a lowering of the low target in RQs. From the verb onwards, the *f0* trajectory diverges, with RQs having lower *f0* values than ISQs. Hence, along with the lexical function of *f0* in Standard Chinese, speakers also use *f0* to mark a difference in illocution type, i.e., the difference between ISQs and RQs. We now turn to prosodic cues other than *f0*, which are duration and voice quality.

3.1.2. Duration

We first tested whether the global utterance duration differed as a function of *illocution type*: Polar-RQs were 163 ms longer than polar-ISQs (1498 ms vs 1335 ms; ß = 0.165; *SE* = 0.03, *df* = 20.50, *t* = 5.42, $p_{adj} < 0.0001$) and *wh*-RQs were 166 ms longer than *wh*-ISQs (1280 ms vs 1114 ms; ß = 0.172; *SE* = 0.02, *df* = 175.2, *t* = 7.75, $p_{adj} < 0.0001$). Since *illocution type* and *question type* did not interact ($p_{adj} = 0.91$), we assume that durational differences between ISQs and RQs hold independently of question type. For exploratory purposes, we assessed whether the longer duration in RQs was carried by a specific part in the utterance (Lo & Kiss, 2020; or whether the lengthening applied globally). To this end, we tested for an interaction between *illocution type* and *word* in an omnibus model that combined polar and *wh*-questions.

[8] The R syntax of the final GAMM was the following: f0.gamB_acf_scat_polar=bam(f0 ~ illocution_type+s(Normtime, by=illocution_type, bs = 'tp', k = 15) +s(Normtime, vp, bs='fs', m = 1) + s(Normtime, item, bs='fs', m = 1), data=data, rho = rhoval, AR.start = data $start_event, method="fREML", discrete=T, family="scat").

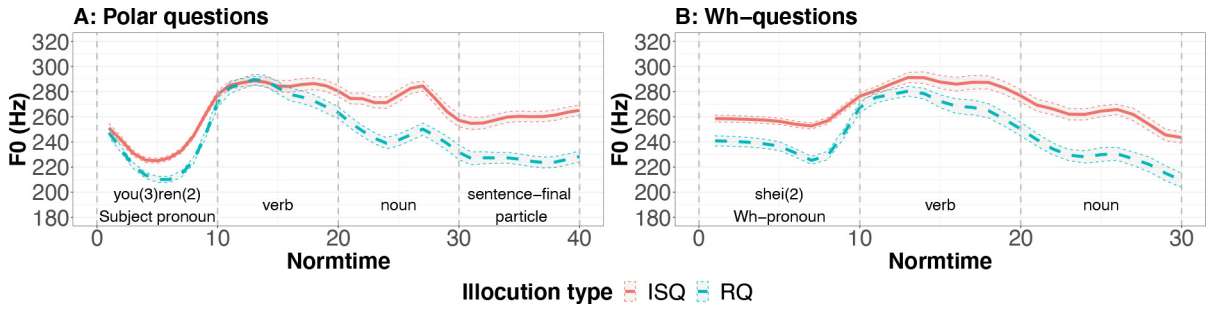

Fig. 2. Time-normalized average *f0* trajectory for the target questions in two illocution types (ISQ in red (solid line) vs RQ in blue (dashed line)) for polar (left, A) and *wh*-questions (right, B). The x-axis shows the normalized time. Ten equidistant *f0* measures per word were extracted from sonorant parts of the words with corrected *f0* using the Praat script Prosody Pro (Y. Xu, 2013); non-sonorant parts were interpolated.

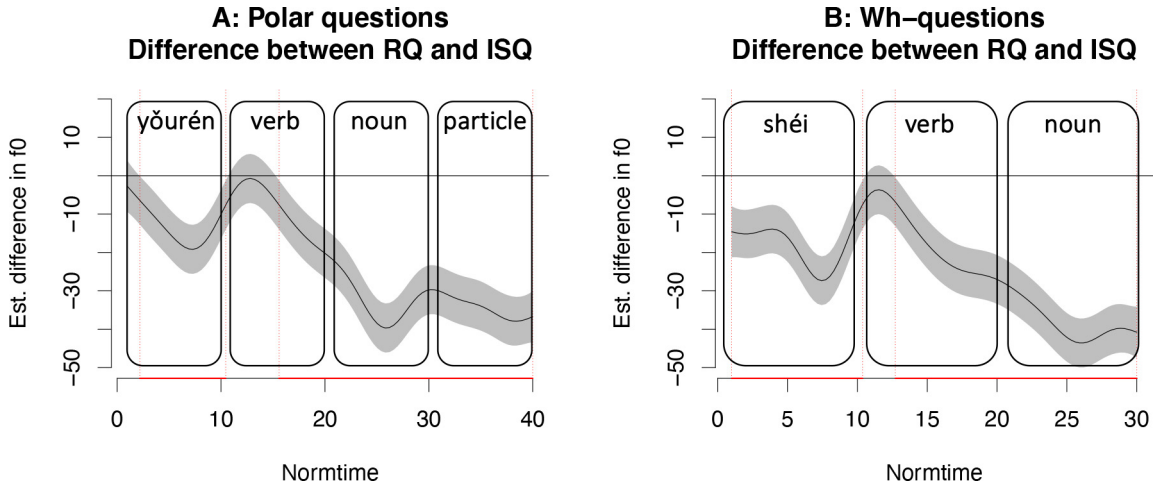

Fig. 3. Predicted difference in *f0* (RQ minus ISQ) for polar questions (left panel) and *wh*-questions (right panel). Values below 0 indicate that RQs are lower in *f0*, values above zero the reverse. The dark grey shading displays the 95%CI (confidence interval) of the predicted mean difference. The difference in *f0* becomes significant if zero is not included in the 95%CI. Significant areas are delimited by the vertical red lines.

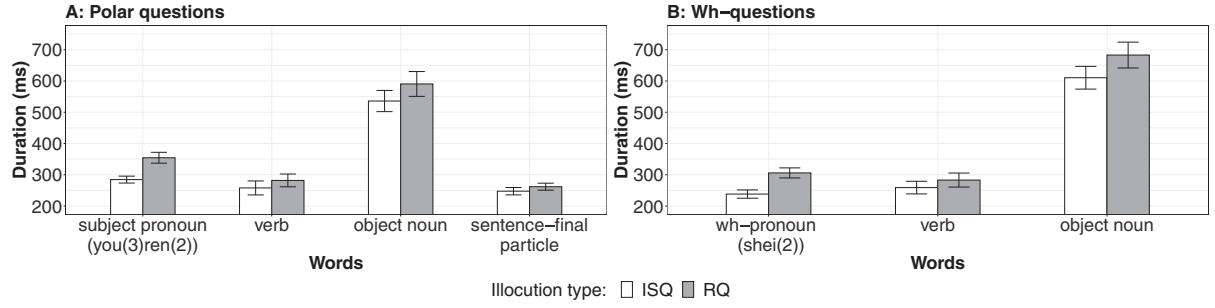

Fig. 4. Average word durations in polar questions (A, left) and *wh*-questions (B, right) in the two illocution types. ISQs are presented in white, RQs in grey. Error bars represent the standard error of the mean.

The variable *word* included the individual words in the target questions (*yǒurén* or *shéi*, verb, object noun). The interaction between *illocution type* and *word* was significant (χ^2 = 9.03, *df* = 2, p_{adj} < 0.05). Hence, the durational differences between ISQs and RQs affected the individual parts of the utterance differently, see Fig. 4. In particular, in both question types, the relative difference across illocution type was strongest for the first word, with *yǒurén* 'anyone' being 25% longer in polar-RQs than in polar-ISQs, and *shéi* 'who' being 28% longer in *wh*-RQs than *wh*-ISQs. The proportional difference between illocution types in the other words for RQs was less than 12%: verb: 9% in both polar and *wh*-questions, object noun: 10% in polar and 12% in *wh*-questions, sentence-final particle *me* in polar question: 5%. All words except for the sentence-final particle (p_{adj} = 0.26) were longer in RQs than in ISQs (all other p_{adj} < 0.05). To sum up, RQs are generally longer than ISQs, with the largest difference between illocution types occurring for the pronominal subject *yǒurén* in polar questions and the *wh*-word *shéi* in *wh*-questions.

3.1.3. Voice quality

The majority of instances at the different measure points for voice quality showed modal voice (83% of the overall labels were modal voice, 17% were glottalized voice; breathy voice

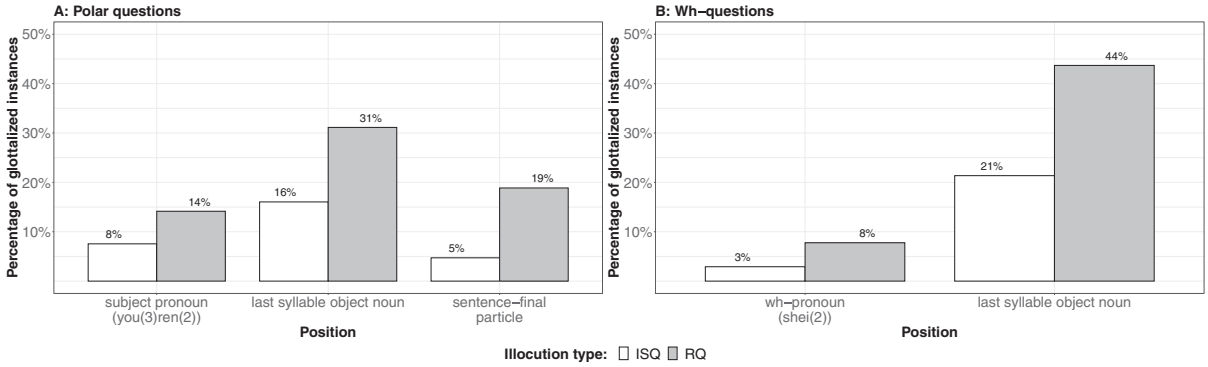

Fig. 5. Proportion of glottalized instances at different positions in the question (A for polar and B for *wh*-questions). ISQs are presented in white, RQs in grey. Proportions relate to the total number of instances in each condition (bar), i.e., *N* = 106 polar-ISQs, *N* = 106 polar-RQs, *N* = 103 *wh*-ISQs, *N* = 103 *wh*-RQs. The respective other realizations are modal voice; breathy voice did not occur.

did not occur at all in our data). Fig. 5 shows the proportion of glottalized voice for polar (A, left) and *wh*-questions (B, right) for ISQs and RQs – at the different measurement points across the question (first word, final syllable in object noun, and sentence-final particle for polar questions); the respective other realizations are modal voice. Note that proportions relate to the total number of instances in each condition (bar), i.e., *N* = 106 polar-ISQs, *N* = 106 polar-RQs, *N* = 103 *wh*-ISQs, *N* = 103 *wh*-RQs. *Position* (i.e., first word, final syllable in object noun, and sentence-final particle in polar question) and *illocution type* did not interact (p_{adj} = 0.30), as revealed in a combined model for polar and *wh*-questions, suggesting no evidence to assume that the distinction in voice quality between ISQs and RQs was different for the different positions in the utterance. There was no interaction between *illocution type* and *question type* (p_{adj} = 0.93), corroborating that the difference between ISQs and RQs in terms of voice quality marking was comparable in both question types. For both polar and *wh*-questions, there are more instances of glottalized voice in RQs than in ISQs (polar questions: 21.4% vs 9.4%, ß = 1.35, *SE* = 0.28, *z* = 4.9, p_{adj} < 0.0001; *wh*-questions: 25.7% vs 12.1%, ß = 1.59, *SE* = 0.37, *z* = 4.3, p_{adj} < 0.0001).

Taken together, Standard Chinese RQs showed an overall higher number of instances of glottalized voice than ISQs, independently of *question type* and *position*. From previous research, we know that other factors also influence the occurrence of glottalized voice beyond illocution type, e.g., mean *f0*, lexical tone, and position (Chen & Gussenhoven, 2008; Kuang, 2017; Shih, 1997, 2000; Xu, 1999). In a first attempt to quantify the strength of the predictor illocution type for the occurrence of glottalized voice – in relation to these other factors – we used a random forest model. This model was trained to predict the occurrence of glottalized voice in the final syllable of the object noun (the syllable in which all four tones occurred, four times Tone 3, six times all other tones). Depending on question type, this syllable was in utterance-final position (in *wh*-questions) or in the penultimate position (in polar questions due to the sentence-final particle). The model revealed lexical tone to be the most important predictor for voice quality in the last syllable of the noun, followed in importance by speaker, illocution type (ISQ vs RQ), and position, see Appendix B for details. This model puts the strength of the factor illocution type into perspective. Specifically, ISQs and RQs are distinguished by voice quality marking in Standard Chinese, but the occurrence of glottalization is intertwined with lexical tone and also depends on the speaker.

3.2. Interplay between individual cues to RQs

The acoustic analyses in Section 3.1 showed that RQs are realized with an overall lower *f0* trajectory (mainly due to a lowering of the low tones in the first word and a compressed *f0* range towards the end of the utterance), longer durations (overall, but especially for the first word), and more instances of glottalized voice (both at the beginning and the end of the utterance). These acoustic comparisons between illocution type (ISQ vs RQ) identify individual cues to RQs, but they do not provide information about the interplay between these cues, i.e., whether these cues are modified together or whether one cue might substitute the other. To pursue Research Question 2, which addresses exactly this interplay between the cues, we analysed whether speakers used duration and mean *f0* in tandem for each illocutionary pair (ISQ and corresponding RQ, *N* = 200). This was operationalized by calculating the difference in mean *f0* and in utterance duration between the RQ and ISQ production of each pair (Δduration and Δf0, henceforth). Each illocutionary pair was further coded with respect to whether or not it was marked for voice quality. Recall that 'yes' was assigned if and only if one of the positions labelled for voice quality (first word, last syllable in object noun, plus sentence-final particle in polar questions) had modal voice in the ISQ version and glottalized voice in the corresponding RQ *and* none of the positions in the respective illocutionary pair showed the reversed contrast (i.e., glottalized in the ISQ and modal voice in the RQ version). Otherwise, this illocutionary pair was coded as 'no'. Fig. 6 shows Δf0 (in Hz, ISQ minus RQ) against Δduration (in seconds, RQ minus ISQ). Each dot represents one illocutionary pair. Illocutionary pairs that are marked by voice quality are plotted in orange triangles (voice quality marking: yes), those that are not have black dots. Regression lines are shown for illocutionary pairs with (orange solid line) and without voice quality marking (black dotted line); the standard error of the regression line is shown in grey shading.

Overall, there was a moderate positive correlation between Δf0 and Δduration for polar questions (*r* = 0.50 [95%CI: 0.34; 0.64], *t* = 5.87, *df* = 101, p_{adj} < 0.0001) and a weak to moderate positive correlation for *wh*-questions (*r* = 0.41 [95%CI: 0.23;

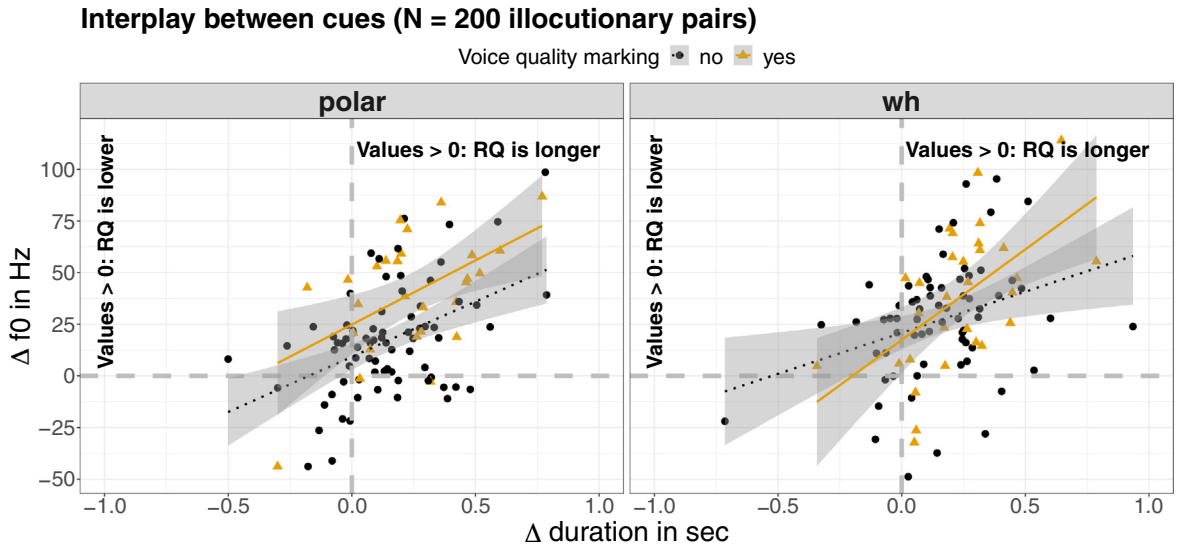

Fig. 6. Scatterplot for the relationship between the difference in the duration of the utterance and the difference in mean *f0* between an ISQ and the corresponding RQ ($N = 200$). Each dot/triangle indicates a question pair (ISQ and corresponding RQ); Pairs with voice quality marking are represented by an orange triangle, pairs with no marking by black dots. The regression line for pairs with voice quality marking is solid orange, for pairs that show no marking it is dotted black.

0.56], $t = 4.40$, $df = 95$, $p_{adj} < 0.0001$). This shows that mean *f0* and duration tend to be modified together, rather than employed in a compensatory way. Regarding voice quality, we observe that for polar questions, voice quality marking is more likely for larger Δf0 (orange solid regression line above black dotted regression line), while for *wh*-questions, we find a steeper slope of the regression line for those illocutionary pairs with voice quality marking, indicating that these pairs even show larger Δf0 at similar Δduration. It is obvious for both question types, however, that illocutionary pairs marked by a voice quality contrast (orange triangles in Fig. 6) are spread out along the regression line. This indicates that voice quality marking is not restricted to illocutionary pairs that show weak prosodic marking otherwise (towards the origin of coordinate system in Fig. 6), but it also occurs in pairs that are more strongly marked by *f0* and duration (towards the right upper corner in Fig. 6). Hence, despite the differences between the two question types (regarding syntax, semantics, and here also number of syllables), the prosodic marking of illocution type is strikingly similar: Mean *f0* and duration marking show a positive relationship, indicating that the cues are used together; voice quality marking also very frequently co-occurs with the marking of *f0* and duration. On the individual level, eight out of ten participants show the positive relation between Δf0 and Δduration, while for two speakers there is no or a slightly negative relation between the two cues (see individual analysis in supplementary analysis script).

Summarising our findings on the interplay between cues to RQs in Standard Chinese, our data do not suggest a compensatory cue trading between *f0*, which is considerably constrained by lexical tone, and the other prosodic cues duration and voice quality. On the contrary, our data show that Standard Chinese speakers jointly – and consistently – use *f0* and duration as cues to RQs. Voice quality seems to be a subordinate or additional cue, which depends on the speaker and, most importantly, on the lexical tone (cf. random forest analysis, in Section 3.1.3).

4. General discussion

Our findings showed that Standard Chinese RQs, as compared to ISQs, are realized with lower *f0*, longer duration and more instances of non-modal (glottalized) voice quality, both in polar and in *wh*-questions, supporting H1. Hence, the prosodic differences between ISQs and RQs are not confined to a particular question type. We have thus shown that essentially the same prosodic cues are used in the tone language Standard Chinese that have also been observed for intonation languages such as English and German, among others (see Section 1.1 above). Given the primary lexical function of *f0* in tone languages, which it does not have in intonation languages, the similarities between the two types of languages are particularly noteworthy. With regard to the interplay between cues to RQs, we find that in most illocutionary pairs, *f0* is modified in tandem with duration and voice quality (cf. Q2). In the remainder of this section, we will first discuss the implications arising from the findings on the prosodic differences in Standard Chinese ISQs vs RQs (Section 4.1). From a cross-linguistic perspective, we will also elaborate on the common use of *f0*, duration and voice quality as prosodic cues to RQs in typologically different languages, and potential explanations (assertive force, focus, speaker attitude) that unite the present findings on RQs in Standard Chinese with the findings on RQs in other, non-tonal languages. In Section 4.2, we address the implications of the interplay between prosodic cues to RQs for the modelling of the relation between prosody and meaning.

4.1. Cues to RQs in Standard Chinese and other (typologically different) languages

This section is organized according to cues, starting with ***f0***. RQs were realized with lower mean *f0* than ISQs in Standard Chinese polar and *wh*-questions, in line with previous findings that Cantonese and Standard Chinese *wh*-RQs have lower *f0*

associated with the sentence-final particle than corresponding *wh*-ISQs (Lo & Kiss, 2020; Lo et al., 2019a). Lower mean *f0* is not restricted to tone languages but also occurs in intonation languages (Beyssade & Delais-Roussarie, 2022 for French; Sahkai et al., 2022 for Estonian; other studies on intonation languages have focused on the differences in pitch accents and boundary tones rather than on global features of *f0* to arrive at a more detailed and language-specific analysis). The *f0* distinction between ISQs and RQs in Standard Chinese also fits in with previous research on pitch modifications for the purpose of marking affect in Standard Chinese (Li et al., 2011), as well as on prosodic characteristics of interrogatives as opposed to declaratives (Lee, 2005; Liu & Xu, 2005; Liu et al., 2016; Yang, 2018; Yang, Gryllia, Pablos, & Cheng, 2019; Yuan, 2006). Given that RQs have functionally been considered assertion-like (cf. Han, 2002, p. 202) and that assertions are typically realized as declaratives in terms of syntactic structure, it is not surprising that RQs prosodically resemble declaratives in that they display lower *f0* compared to information-seeking questions (for Standard Chinese: Lee, 2005; Liu & Xu, 2005; Liu et al., 2016; Yang, 2018; Yang et al., 2019; Yuan, 2006). Conversely, higher pitch, both globally and at specific positions in the utterance (here final) have been associated with inquisitive utterances (information-seeking questions) as compared to statements in a variety of languages (cf. Hirst & Di Cristo, 1998, pp. 24-26 for overview).

The *f0* modulation we observe cannot solely be explained by the marking of **assertive force**, otherwise we would have observed a difference in register only (e.g., RQs uniformly lower than ISQs). Yet, the *f0* trajectories of ISQs vs RQs paint a more detailed picture that goes beyond a register difference. There is an expansion of the *f0* excursion for the pronominal subject *yǒurén* 'anyone' in polar questions and the *wh*-word *shéi* 'who' in *wh*-questions in RQs as compared to ISQs. From the second word onwards (i.e., the verb), the *f0* trajectories diverged, with RQs becoming increasingly lower than ISQs. Hence, it seems that speakers increased the *f0* range to mark an interrogative as rhetorical in the beginning of the utterance, and after that reduced the *f0* range in RQs as compared to ISQs (cf. Yuan, 2006, who shows the distinction in *f0* between statements and polar questions to become larger towards the end of the interrogative). This observation of an increased *f0* range would be compatible with **prosodic focus marking** on the first word in the RQs, followed by post-focal compression (Gårding et al., 1983; Jin, 1996; Xu, 1999; Xu & Xu, 2005, see Chen, 2022 for overview). Such an interpretation is in line with the fact that the first word was also lengthened most strongly (in both polar and *wh*-questions), further increasing their prominence. A post-hoc prominence rating task by a native speaker of Standard Chinese, who indicated whether the first word (*yǒurén* 'anyone' in polar questions and *shéi* 'who' in *wh*-questions) or another part of the question sounded most prominent to her, showed that RQs were often perceived as having the main sentence prominence on the first word (37% in polar and 56% in *wh*-questions); the main prominence for ISQs was perceived on the sentence-final object noun in almost all cases (100% in polar and 96% in *wh*-questions). Interestingly, Dehé and Braun (2020b) observe a similar shift in nuclear accent position for English polar questions: About 25% of the RQs in their data set were realized with the nuclear accent on the subject pronoun 'anyone' and no accent on the sentence-final noun ("Does ANYONE eat lemons?", capitals indicate the word that carries the nuclear pitch accent). This pattern never occurred for ISQs. For polar questions, a focus on *anyone* can turn the indefinite subject pronoun into a negative polarity item, which is only compatible with RQs and not with ISQs (e.g., Han, 2002). Likewise, a focused *wh*-word may strengthen the salience of the empty set interpretation intended by the RQs ("Nobody likes lemons").

Our data-driven idea of focus being realized on the first word in RQs, but not in ISQs, however, challenges the semantic assumption that the *wh*-word is considered to have focus in neutral *wh*-questions (e.g., Deguchi & Kitagawa, 2002; Lambrecht & Michaelis, 1998; Yang et al., 2020). The theoretical question that arises is whether – based on our empirical data – we would still have to assume focus in *wh*-ISQs on the *wh*-word, and how to theoretically differentiate the two types of *wh*-questions (ISQ vs RQ) if both have focus on the *wh*-word, but perceptually, the prominence location differs. One possibility might be that in ISQs, we observe a misalignment between semantic focus and its prosodic manifestation, i.e., semantic focus on the *wh*-word, prosodic manifestation in sentence-final position (see Chen, 2006, for an example of durational marking of corrective focus not being self-contained on the focused element, but spilling over to the following syllables; cf. Rooth, 2008, on scope of focus). It might also be the case that the prosodic realization of focus in RQs is more salient than that in ISQs, which might have influenced the native speaker in the prominence rating task towards more 'first word' ratings in RQs as compared to ISQs. In that case, focus marking might be a matter of different degrees of emphasis between ISQs and RQs, with RQs being more strongly marked than ISQs – along the lines of Chen and Gussenhoven (2008) for different degrees of emphasis in focus marking. There are hence several possibilities that could explain our findings of the expansion of the *f0* excursion for the first word in RQs as compared to ISQs, which would indeed be worthy of further research in the future.

From a cross-linguistic perspective, Standard Chinese resembles other, typologically different languages in that it also uses *f0* to mark a question as rhetorical. Even though the implementation shows language-specific aspects (i.e., certain pitch accent types or edge tones are favoured in RQs in intonation languages, cf. Dehé et al., 2022), there seem to be cross-linguistic similarities with respect to the position of sentence accent (Dehé & Braun, 2020b).

Duration. RQs in Standard Chinese were produced with longer overall duration, in line with findings in a variety of typologically different languages (Beyssade & Delais-Roussarie, 2022 for French; Braun et al., 2019 for German; Dehé & Braun, 2020a for Icelandic; Dehé & Braun, 2020b for English; Dehé et al., 2018 for Icelandic; Lo et al., 2019b on Cantonese; Miura & Hara, 1995 for Japanese; Sahkai et al., 2022 for Estonian; Sorianello, 2018 for Italian). Longer duration hence seems to be a stable characteristic of RQs across languages, both occurring in lab-like settings and in RQs in spontaneous productions (Braun et al., 2020). This again ties in with durational differences in other speech acts, such as the distinction between statements and interrogatives in Standard Chinese, both for polar and *wh*-questions (X. Liu et al., 2016; Yang,

2018; Yuan, 2006). Also, faster speaking rate has been observed in declarative questions as compared to declaratives (Niebuhr et al., 2010, for German; van Heuven & van Zanten, 2005, for Manado Malay, Orkney English, and Dutch; and for exclamatives vs ISQs for German, Wochner, 2022). We can hence generalize that utterances with an **assertive force** (statements and rhetorical questions) are longer than genuine information-seeking questions, which lack assertive force. Given that similar prosodic differences – both for *f0*-related and durational cues – have been reported for the distinction between statements and questions in Standard Chinese (Lee, 2005; Liu & Xu, 2005; Liu et al., 2016; Yang, 2018; Yang et al., 2019; Yuan, 2006), a logical next step in future studies will be to compare Standard Chinese RQs to string-identical assertions (Wochner, 2022 on German ISQs, RQs, and assertions).

Voice quality. We find that Standard Chinese RQs are more often realized with glottalized voice than string-identical ISQs (see also Lo & Kiss, 2020, on creaky voice in sentence-final particles in wh-RQs). This was the case for both the first word, the final syllable of the noun and the sentence-final particle, although glottalization was generally more frequent towards the end of the utterance. Similarly, voice quality differentiates between ISQs and RQs in a number of languages (Braun et al., 2019; Dehé & Braun, 2020b; Dehé & Wochner, 2022; Sahkai et al., 2022). While the voice quality contrast is spelled out differently across languages, either as a contrast between modal and breathy (English, German, Icelandic) or between modal and glottal (Standard Chinese, Estonian), the generalization is that RQs are more often produced with non-modal voice quality than ISQs. In Standard Chinese, the presence of glottalization was influenced more by lexical tone identity (and hence *f0*) and speaker identity than by illocution type (cf. random forest), which suggests that glottalization is not a strictly independent cue. Breathy voice, a cue to RQs in other languages (Braun et al., 2019; Dehé & Braun, 2020b; Dehé & Wochner, 2022), might be more independent of *f0*. Future research needs to further study the relation between voice quality and *f0* in the marking of illocution type in other languages.

Given that the possible answers to the RQs in our study were negative, e.g., *nobody eats lemons*, it might be the case that speakers additionally convey a **negative attitude** when asking an RQ. Conceivably, stronger attitudes may have led to a stronger marking of illocution type (cf. a post-hoc analysis of Neitsch, 2019, for the influence of attitude on the prosodic marking of RQs in German). For instance, lower *f0*, longer duration and glottalized voice have also been reported to mark disgust, a negative attitude, in Standard Chinese (Li et al., 2011; Liu & Pell, 2012; Yuan et al., 2002). As indicated earlier, a pre-test confirmed that our materials were on average ambiguous as to whether participants liked them or not. Clearly though, for individual items there might be differences in the strength of attitude, which may have interacted with the marking of illocution type.[9]

4.2. Interplay between cues to RQs in Standard Chinese

In response to our second research question, we analysed the interplay between the different cues to illocution type. From other languages and linguistic phenomena it is known that prosodic cues may either have trading relations (with one cue compensating for the other, e.g., Kim, 2020; Niebuhr, D'Imperio, Gili Fivela, & Cangemi, 2011; cf. Schertz & Clare, 2019) or, conversely, prosodic cues may be used in tandem (Braun, 2006; Kim, 2020). For Standard Chinese, as a tone language, we considered the possibility that speakers use fewer *f0* modifications and instead make more use of duration and voice quality. Our findings revealed that this is not the case and that *f0* and duration are largely modified in tandem, as evidenced by a positive correlation between the differences in mean *f0* and sentence duration across illocutionary pairs. Hence, the stronger the marking in terms of *f0*, the stronger the durational difference; if voice quality distinctions occurred, they did so in addition to other cues.

What does this joint modification of cues in the encoding of illocution type imply for the relationship between prosody and meaning? Recently, so-called 'prosodic constructions' have been suggested in the literature (Ward, 2019; Ward & Gallardo, 2017) in order to explain how prosodic cues combine to form meaningful configurations or constructions. Prosodic constructions are defined as "recurring temporal patterns of prosodic activity that express specific meanings and functions and which typically involve not only pitch contours but also energy, rate, timing and articulation properties" (Ward & Gallardo, 2017, 3f., but see Burdin, 2020; Huttenlauch, Feldhausen, & Braun, 2018, for arguments against prosodic constructions). The finding of joint modification of *f0*, duration, and voice quality is compatible with a prosodic construction approach. However, given that these cues are also jointly employed to encode focus (*f0*, duration), interrogativity (*f0*, duration), and emotions/attitudes (*f0*, duration, and voice quality), considering the observed pattern 'prosodic constructions' does not add explanatory value. Based on our findings, we hence argue that a bundle of cues is used to convey rhetorical illocution, just like for emphatic focus (Chen & Gussenhoven, 2008), but we refrain from confining the present combinations of cues to RQs alone. The weighting of cues to RQ interpretation needs to be further investigated in perception studies (see Kharaman, Xu, Eulitz, & Braun, 2019 on German; Miura & Hara, 1995 on Japanese). Such experiments assess how speakers weigh cues to identify and distinguish RQs from ISQs. Kharaman et al. (2019), for instance, orthogonally crossed nuclear pitch accent (late- vs early-peak accent), duration (short vs long), and voice quality (breathy vs modal) and asked German listeners to judge target interrogatives as either ISQs or RQs. While the interplay of all cues (late-peak nuclear accent, long duration, breathy voice) led to 97% of RQ responses, the presence of two cues still resulted in more than 75% of RQ identifications (80% for late-peak and breathy voice, and 75% for late-peak and long duration). We will pursue this issue for Standard Chinese in future research, with a particular focus on the interaction between lexical tone and intonation.

5. Conclusion

To conclude, our study reveals that speakers of Standard Chinese jointly employ lower *f0*, longer duration, and – less consistently more glottalized voice quality – to prosodically mark the illocutionary force of a question as rhetorical. Impor-

[9] It was impossible to control for the strength of attitude (based on pre-test ratings) in the statistical models because of collinearity with other predictors.

tantly, these prosodic differences are not confined to a particular question type, but equally apply to both polar and *wh*-questions. Our findings further reveal that *f0* and duration are largely modified in tandem, while voice quality seems to be an additional cue. From a cross-linguistic perspective, we conclude that *f0* and duration modifications are robust cues to rhetorical illocution, with non-modal voice quality being a more optional choice (at least in the languages tested). As shown above, there are three possible explanations according to which cross-linguistic differences can be united: (A) The marking of assertive force (duration and *f0*), (B) the marking of focus (accent position, expanded *f0* range, longer duration), and (C) the marking of speaker attitude (lower *f0*, non-modal voice quality). Future work needs to find ways to check whether and how these possible mechanisms can be disentangled.

Funding

The research was funded by the DFG as part of research unit 'Questions at the Interfaces' (FOR 2111, P6), Grant No BR 3428/4-1,2 and DE 876/3-1,2.

Author contributions

Bettina Braun and Nicole Dehé are Co-PIs on the funded project on rhetorical questions mentioned in the Section *Funding*. They designed the experiment and supervised testing. Katharina Zahner-Ritter annotated the data, prepared the data for analysis, conducted the acoustic and statistical analyses and drafted the manuscript. Yiya Chen advised on language-specific aspects in the acoustic analysis and interpretation and conceptualization. All authors jointly discussed the experimental results their implications, and are responsible for the final version of the manuscript.

CRediT authorship contribution statement

Katharina Zahner-Ritter: Conceptualization, Methodology, Formal analysis, Data curation, Visualization, Writing – original draft. **Yiya Chen:** Conceptualization, Writing – review & editing. **Nicole Dehé:** Conceptualization, Investigation, Methodology, Writing – review & editing, Funding acquisition. **Bettina Braun:** Conceptualization, Investigation, Methodology, Writing – review & editing, Funding acquisition.

Acknowledgements

We are grateful to Manluolan Xu for translation of the experimental stimuli, testing the participants, and help with annotation, to Meiqiao Li and Tianyi Zhao for help with annotation and for native advice, and also to Jasmin Rimpler for help with data preparation. We also thank Cesko Voeten for helpful practical advice on GAMM modelling, and Haiping Long, Tian Li, Jun Liu, Mengzhu Yan and the audiences at P&P 2019 in Düsseldorf and virtual Speech Prosody 2020 in Tokyo for discussion and feedback. Finally, we thank the reviewers and editor for their helpful comments.

Appendix A

List of all target questions (polar and *wh*-version) used in the experiment. The last tone in the sentence-final noun is additionally indicated in the final column (items are grouped by the last tone in the sentence-final noun, 6 × Tone 1, 6 × Tone 2, 4 × Tone 3, 6 × Tone 4). In total, we made six adjustments compared to the German materials in previous work (cf. Braun et al., 2019) for cultural reasons. For instance, the item *angora* (安哥拉 [āngēlā]) was replaced by *sable fur* (貂(diāo)皮(pí)) as this was judged by native informants to be better known in Chinese culture.

List of all target questions (polar and *wh*-version) used in the experiment.

wh-questions	polar questions	Last tone
谁(shéi)吃(chī)小(xiǎo)虾(xiā)? 'Who eats shrimps?'	有(yǒu)人(rén)吃(chī)小(xiǎo)虾(xiā) 么(ma/me)? 'Does anyone eat shrimps?'	Tone1
谁(shéi)喜(xǐ)欢(huān)肝(gān)? 'Who likes liver?'	有(yǒu)人(rén)喜(xǐ)欢(huān)肝(gān)么(ma/me)? 'Does anyone like liver?'	Tone1
谁(shéi)送(sòng)百(bǎi)合(hé)花(huā)? 'Who gives lilies (as a present)?'	有(yǒu)人(rén)送(sòng)百(bǎi)合(hé)花(huā)么(ma/me)? 'Does anyone give lilies (as a present)?'	Tone1
谁(shéi)吃(chī)苦(kǔ)瓜(guā)? 'Who eats bitter melons?'	有(yǒu)人(rén)吃(chī)苦(kǔ)瓜(guā) 么(ma/me)? 'Does anyone eat bitter melons?'	Tone1
谁(shéi)读(dú)长(cháng)篇(piān)小(xiǎo)说(shuō)? 'Who reads novels?'	有(yǒu)人(rén)读(dú)长(cháng)篇(piān)小(xiǎo)说(shuō)么(ma/me)? 'Does anyone read novels?'	Tone1
谁(shéi)想(xiǎng)要(yào)玫(méi)瑰(guī)? 'Who would like roses?'	有(yǒu)人(rén)想(xiǎng)要(yào)玫(méi)瑰(guī)么(ma/me)? 'Does anyone want roses?'	Tone1
谁(shéi)穿(chuān)貂(diāo)皮(pí)? 'Who wears sable fur?'	有(yǒu)人(rén)穿(chuān)貂(diāo)皮(pí)么(ma/me)? 'Does anyone wear sable fur?'	Tone2
谁(shéi) 想(xiǎng)喝(hē)菊(jú)花(huā)茶(chá)? 'Who wants camomile?'	有(yǒu)人(rén)想(xiǎng)喝(hē)菊(jú)花(huā)茶(chá)么(ma/me)? 'Does anyone want camomile?'	Tone2

Appendix A. (*continued*)

wh-questions	polar questions	Last tone
谁(shéi)喜(xǐ)欢(huān)鱼(yú)肝(gān)油(yóu)? 'Who likes cod-liver oil?'	有(yǒu)人(rén)喜(xǐ)欢(huān)鱼(yú)肝(gān)油(yóu)么(ma/me)? 'Does anyone like cod-liver oil?'	Tone2
谁(shéi) 吃(chī)榴(liú)莲(lián)? 'Who eats durian?'	有(yǒu)人(rén)吃(chī)榴(liú)莲(lián)么(ma/me)? 'Does anyone eat durian?'	Tone2
谁(shéi)吃(chī)柠(níng)檬(méng)? 'Who eats limes?'	有(yǒu)人(rén)吃(chī)柠(níng)檬(méng)么(ma/me)? 'Does anyone eat limes?'	Tone2
谁(shéi)培(péi)育(yù)蠕(rú)虫(chóng)? 'Who breeds worms?'	有(yǒu)人(rén)培(péi)育(yù)蠕(rú)虫(chóng)么(ma/me)? 'Does anyone breed worms?'	Tone2
谁(shéi)跳(tiào)霹(pī)雳(lì)舞(wǔ)? 'Who dances breakdance?'	有(yǒu)人(rén)跳(tiào)霹(pī)雳(lì)舞(wǔ)么(ma/me)? 'Does anyone dance breakdance?'	Tone3
谁(shéi)想(xiǎng)去(qù)博(bó)物(wù)馆(guǎn)? 'Who wants to go to the museum?'	有(yǒu)人(rén)想(xiǎng)去(qù)博(bó)物(wù)馆(guǎn)么(ma/me)? 'Does anyone want to go to the museum?'	Tone3
谁(shéi)知(zhī)道(dào)鱼(yú)腥(xīng)草(cǎo)? 'Who knows Houttuynia?'	有(yǒu)人(rén)知(zhī)道(dào)鱼(yú)腥(xīng)草(cǎo)么(ma/me)? 'Does anyone know Houttuynia?'	Tone3
谁(shéi)需(xū)要(yào)模(mú)板(bǎn)? 'Who needs stencils?'	有(yǒu)人(rén)需(xū)要(yào)模(mú)板(bǎn)么(me)? 'Does anyone need stencils?'	Tone3
谁(shéi)学(xué)代(dài)数(shù)? 'Who studies algebra?'	有(yǒu)人(rén)学(xué)代(dài)数(shù)么(ma/me)? 'Does anyone study algebra?'	Tone4
谁(shéi)吃(chī)番(fān)茄(qié)肉(ròu)酱(jiàng)面(miàn)? 'Who eats pasta Bolognese?'	有(yǒu)人(rén)吃(chī)番(fān)茄(qié)肉(ròu)酱(jiàng)面(miàn)么(ma/me)? 'Does anyone eat pasta Bolognese?'	Tone4
谁(shéi)吃(chī) 内(nèi)脏(zàng)? 'Who eats innards?'	有(yǒu)人(rén)吃(chī)内(nèi)脏(zàng)么(ma/me)? 'Does anyone eat innards?'	Tone4
谁(shéi)喜(xǐ)欢(huān)蛋(dàn)黄(huáng)酱(jiàng)? 'Who likes mayonnaise?'	有(yǒu)人(rén)喜(xǐ)欢(huān)蛋(dàn)黄(huáng)酱(jiàng)么(ma/me)? 'Does anyone like mayonnaise?'	Tone4
谁(shéi)读(dú)人(rén)物(wù)传(zhuàn)记(jì)? 'Who reads biographies?'	有(yǒu)人(rén)读(dú)人(rén)物(wù)传(zhuàn)记(jì)么(ma/me)? 'Does anyone read biographies?'	Tone4
谁(shéi) 喜(xǐ)欢(huān)芹(qín)菜(cài)? 'Who likes celery?'	有(yǒu)人(rén) 喜(xǐ)欢(huān)芹(qín)菜(cài)么(ma/me)? 'Does anyone like celery?'	Tone4

Appendix B. Random forest: Predicting voice quality in final syllable of the noun

Using the R-package *randomForest* (Liaw & Wiener, 2002), we fitted a random forest model to predict the voice quality label (modal vs glottal) for the final syllable in the sentence-final noun as a function of a number of variables, including *illocution type*, *lexical tone* in the last syllable of the sentence-final particle, *speaker*, and *position* (the second to last syllable of the interrogative in polar questions, and the last syllable of the interrogative in *wh*-questions). To train the random forest, we randomly selected 80% for training and 20% for test. The number of trees was set to 1000. Random forests extract the importance of the individual variables using the Gini-index (Liaw & Wiener, 2002), see Fig. B.1. Note that higher values indicate greater importance of a variable. Hence, our model revealed the *lexical tone* in the last syllable of the noun to be most important in predicting the occurrence of voice quality labels at the last syllable of the noun, followed in importance by *speaker*, *illocution type*, and *position*. The accuracy for the random forest models on 20% of unseen data (training set) was 82.1%.

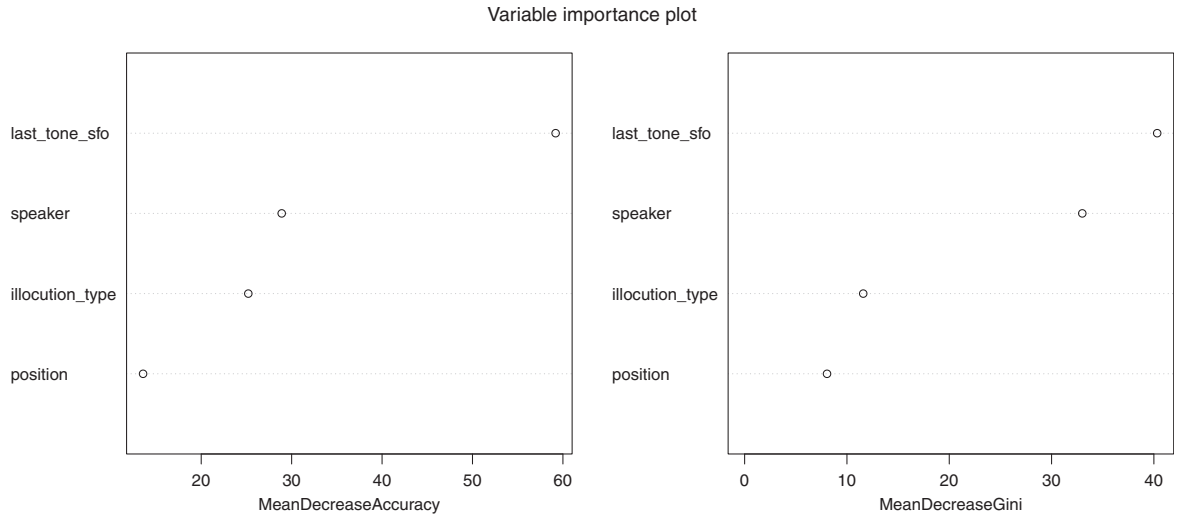

Fig. B.1. Results of the random forest for the voice quality label at the final syllable in the sentence-final noun. 1. The Mean Decrease Gini plot (right panel) reveals the importance of the predictors (last tone in sentence-final noun > speaker > illocution type > position); the Mean Deacrease Accuracy plot shows how much of the accuracy the model loses when excluding a variable (the higher the loss, the more important the variable).

Appendix C. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.wocn.2022.101190.

References

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*, 390–412.

Baayen, R. H., van Rij, J., de Cat, C., & Wood, S. N. (2018). Autocorrelated errors in experimental data in the language sciences: Some solutions offered by Generalized Additive Mixed Models. In D. Speelman, K. Heylen, & D. Geeraerts (Eds.), *Mixed effects regression models in linguistics* (pp. 49–69). Berlin: Springer.

Bayer, J., & Obenauer, H.-G. (2011). Discourse particles, clause structure, and question types. *The Linguistic Review, 28*, 449–491.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, 57*, 289–300.

Beyssade, C., & Delais-Roussarie, E. (2022). The prosody of French rhetorical questions. *Linguistics Vanguard (Special Issue on non-canonical questions), 8*, 277–286.

Biezma, M., & Rawlins, K. (2017). Rhetorical questions: Severing asking from questioning. *Proceedings of SALT, 27*, 302–322.

Boersma, P., & Weenink, D. (2016). Praat: doing phonetics by computer [Computer program]. Version 6.1.42, retrieved from http://www.praat.org/.

Braun, B. (2006). Phonetics and phonology of thematic contrast in German. *Language and Speech, 49*, 451–493.

Braun, B., Dehé, N., Neitsch, J., Wochner, D., & Zahner, K. (2019). The prosody of rhetorical and information-seeking questions in German. *Language and Speech, 62*, 779–807.

Braun, B., Einfeldt, M., Esposito, G., & Dehé, N. (2020). The prosodic realization of rhetorical and information-questions in German spontaneous speech. *Proceedings of the International Conference on Speech Prosody*, Tokyo, Japan, (pp. 342–346).

Burdin, R. S. (2020). Review: Prosodic patterns in English conversation. *Journal of English Linguistics, 48*, 208–212.

Caponigro, I., & Sprouse, J. (2007). In *Rhetorical questions as questions. In Proceedings of Sinn und Bedeutung* (11, pp. 121–133). Barcelona, Spain: Universitat Pompeu Fabra.

Chang, N.-C.-T. (1975). Tones and intonation in Chengdu Dialect (Szechuan, China). *Phonetica, 2*, 59–85.

Chao, Y. R. (1930). A system of tone-letters. *Le Maître Phonétique, 45*, 24–27.

Chao, Y. R. (1956). Tone, intonation, singsong, chanting, recitative, tonal composition and atonal composition in Chinese. In M. Halle, H. Lunt, H. McLean, & C. V. Schooneveld (Eds.), *For Roman Jakobson: Essays on the occasion of his sixtieth birthday* (pp. 52–59). The Hague, The Netherlands: Mouton Publishers.

Chao, Y. R. (1968). *A grammar of spoken Chinese*. Berkeley: University of California Press.

Chen, Y. (2006). Durational adjustment under corrective focus in Standard Chinese. *Journal of Phonetics, 34*, 176–201.

Chen, Y. (2008). The acoustic realization of vowels of Shanghai Chinese. *Journal of Phonetics, 36*, 629–648.

Chen, Y. (2010). Post-focus f0 compression - Now you see it, now you don't. *Journal of Phonetics, 38*, 517–525.

Chen, Y. (2016). Neutral tone. In R. Sybesma, W. Behr, Y. Gu, Z. Handel, & J. Huang (Eds.), *Encyclopedia of Chinese language and linguistics*. Brill: Leiden.

Chen, Y. (2022). Tone and intonation. In C.-. R. Huang, Y.-H. Lin, I.-H. Chen, & Y.-Y. Hsu (Eds.), *The Cambridge Handbook of Chinese Linguistics* (pp. 336–360). Cambridge: Cambridge University Press.

Chen, Y., & Braun, B. (2006). Prosodic realization in information structure categories in standard Chinese. *Proceedings of the 3rd International Conference on Speech Prosody 3*, Dresden, Germany, 54-57.

Chen, Y., & Gussenhoven, C. (2008). Emphasis and tonal implementation in Standard Chinese. *Journal of Phonetics, 36*, 724–746.

Chen, Y., Lee, P., & Pan, H. (2016). Focus and topic marking in Chinese. In C. Féry & S. Ishihara (Eds.), *Oxford Handbook of Information Structure* (pp. 733–752). Oxford: Oxford University Press.

Cheng, L.-L.-S. (1991). *On the typology of wh-questions*. Cambridge: MIT.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*, 37–46.

Deguchi, M., & Kitagawa, Y. (2002). Prosody and wh-questions. *North East Linguistics Society (NELS), 32*, 73–92.

Dehé, N., & Braun, B. (2020a). The intonation of information seeking and rhetorical questions in Icelandic. *Journal of Germanic Linguistics, 32*, 1–42.

Dehé, N., & Braun, B. (2020b). The prosody of rhetorical questions in English. *English Language and Linguistics, 24*, 607–635.

Dehé, N., Braun, B., Einfeldt, M., Wochner, D., & Zahner-Ritter, K. (2022). The prosody of rhetorical questions: A cross-linguistic view. *Linguistische Berichte*, 3–42.

Dehé, N., Braun, B., & Wochner, D. (2018). The prosody of rhetorical vs. information-seeking questions in Icelandic. *Proceedings of the 9th International Conference on Speech Prosody 2018* Poznań, Poland, 403-407.

Dehé, N., & Wochner, D. (2022). Voice quality and speaking rate in Icelandic rhetorical questions. *Nordic Journal of Linguistics*, 1–10. https://doi.org/10.1017/S0332586522000014.

Dehé, N., Wochner, D., & Einfeldt, M. (2022). The interaction of discourse markers and prosody in rhetorical questions in German. *Journal of Linguistics*, 1–25. https://doi.org/10.1017/S0022226722000299.

Fang, H. (2021). The non-interrogative sentence-final particle ne 呢 in Mandarin. *Linguistics in Amsterdam, 14*, 58–90.

Gamer, M., Lemon, J., Fellows, I., & Singh, P. (2012). irr: various coefficients of interrater reliability and agreement. *R package version 0.84* https://CRAN.R-project.org/package=irr.

Gårding, E., Zhang, J., & Svantesson, J.-O. (1983). A generative model for tone and intonation in Standard Chinese based on data from one speaker. *Lund Working Papers, 25*, 53–65.

Groenendijk, J. A. G., & Stokhof, M. J. B. (1984). *Studies on the semantics of questions and the pragmatics of answers*. University of Amsterdam.

Gussenhoven, C. (2004). *The phonology of tone and intonation*. Cambridge: Cambridge University Press.

Han, C.-H. (2002). Interpreting interrogatives as rhetorical questions. *Lingua, 112*, 201–229.

Hirst, D. J., & Di Cristo, A. (1998). A survey of intonation systems. In D. J. Hirst & A. Di Cristo (Eds.), *Intonation systems: A survey of twenty languages* (pp. 1–44). Cambridge: Cambridge University Press.

Huttenlauch, C., Feldhausen, I., & Braun, B. (2018). The purpose shapes the vocative: Prosodic realisation of Colombian Spanish vocatives. *Journal of the International Phonetic Association, 48*, 33–56.

Jin, S. (1996). *An acoustic study of sentence stress in Mandarin Chinese*. Columbus: OSU.

Keating, P., Garellek, M., & Kreiman, J. (2015). Acoustic properties of different kinds of creaky voice. *Proceedings of the International Congress of Phonetic Sciences*, Glasgow.

Kharaman, M., Xu, M., Eulitz, C., & Braun, B. (2019). The processing of prosodic cues to rhetorical question interpretation: Psycholinguistic and neurolinguistics evidence. *Proceedings of the 20th Annual Conference of the International Speech Communication Association (Interspeech)*, Graz, Austria, 1218-1222.

Kim, J. (2020). *Individual Differences in the Production and Perception of Prosodic Boundaries in American English*. The University of Michigan. PhD thesis.

Krifka, M. (2011). Questions. In K. von Heusinger, C. Maienborn, & P. Portner (Eds.). *Semantics: An international handbook of natural language meaning* (Vol. 2, pp. 1742–1785). Berlin: Mouton de Gruyter.

Kuang, J. J. (2017). Covariation between voice quality and pitch: Revisiting the case of Mandarin creaky voice. *Journal of the Acoustical Society of America, 142*, 1693–1706.

Kuang, J. J. (2018). The influence of tonal categories and prosodic boundaries on the creakiness in Mandarin. *Journal of the Acoustical Society of America, 143*, EL509–EL515.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software, 82*, 1–26.

Lambrecht, K., & Michaelis, L. A. (1998). Sentence accent in information questions: Default and projection. *Linguistics and Philosophy, 21*, 477–544.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*, 159–174.

Laver, J. (1980). *The phonetic description of voice quality* (Vol. 31) Cambridge: Univ. Press.

Lee, O. J. (2005). *The prosody of questions in Beijing Mandarin*. The Ohio State University.

Li, A., Fang, Q., & Dang, J. (2011). Emotional intonation in a tone language: Experimental evidence from Chinese. In *Proceedings of the International Congress for Phonetic Sciences XVII, Hong Kong* (pp. 17–21).

Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News, 2*.

Liing, W.-J. (2014). *How to ask questions in Mandarin Chinese*. CUNY Academic Works: City University of New York.

Lin, Y.-H. (2007). *The sounds of Chinese*. Cambridge: Cambridge University Press.

Liu, F., & Xu, Y. (2005). Parallel encoding of focus and interrogative meaning in Mandarin Intonation. *Phonetica, 62*, 70–87.

Liu, P., & Pell, M. D. (2012). Recognizing vocal emotions in Mandarin Chinese: A validated database of Chinese vocal emotional stimuli. *Behavior Research Methods, 44*, 1042–1051.

Liu, X., Li, A., & Jia, Y. (2016). How does prosody distinguish wh-statement from wh-question? A case study of standard Chinese. *Proceedings of the International Conference on Speech Prosody, 2016*, 1076–1080.

Lo, R.-Y.-H., & Kiss, A. (2020). Durational and pitch marking of rhetorical wh-questions in Mandarin. *Proceedings of the 10 th International Conference on Speech Prosody (Speech Prosody 2020), Tokyo, Japan*.

Lo, R.-Y.-H., Kiss, A., & Tulling, M. (2019a). The prosodic properties of the Cantonese sentence-final particles aa1 and aa3 in rhetorical wh-questions. In *Proceedings of the International Congress of Phonetic Sciences (ICPhS), Melbourne, Australia* (pp. 502–506).

Lo, R.-Y.-H., Kiss, A., & Tulling, M. (2019b). *The prosody of Cantonese information-seeking and negative rhetorical wh-questions*. Vancouver, BC: Paper presented at the Canadian Linguistic Association.

Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H. R., & Bates, D. M. (2017). Balancing type 1 error and power in linear mixed models. *Journal of Memory and Language, 94*, 305–315.

Miura, I., & Hara, N. (1995). Production and perception of rhetorical questions in Osaka Japanese. *Journal of Phonetics, 2*, 291–303.

Neitsch, J. (2019). *Who cares about context and attitude? Prosodic variation in the production and perception of rhetorical questions in German*. Konstanz, Germany: University of Konstanz. PhD thesis.

Niebuhr, O. (2012). At the edge of intonation—the interplay of utterance-final F0 movements and voiceless fricative sounds. *Phonetica, 69*, 7–27.

Niebuhr, O. (2017). On the perception of "segmental intonation": F0 context effects on sibilant identification in German. *EURASIP Journal on Audio, Speech, and Music Processing, 2017*, 1–20.

Niebuhr, O., Bergherr, J., Huth, S., Lill, C., & Neuschulz, J. (2010). Intonationsfragen hintergragt - Die Vielschichtigkeit der prosodischen Unterschiede zwischen Aussage- und Fragesätzen mit deklarativer Syntax. *Zeitschrift für Dialektologie und Linguistik, 77*, 304–346.

Niebuhr, O., D'Imperio, M., Gili Fivela, B., & Cangemi, F. (2011). Are there "shapers" and "aligners"? Individual differences in signalling pitch accent category. *Proceedings of the 17th International Congress of Phonetic Sciences, ICPhS*, Hong Kong, China.

Porretta, V., Tucker, B. V., & Järvikivi, J. (2016). The influence of gradient foreign accentedness and listener experience on word recognition. *Journal of Phonetics, 58*, 1–21.

Powell, M. J. D. (2009). The BOBYQA algorithm for bound constrained optimization without derivatives. Technical Report, Cambridge NA Rep. NA2009/06, Univ. Cambridge, Cambridge, pp. 26–46.

Presentation, N. S. (2000). Berkeley, CA.

R Development Core Team, R. (2015). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.

Rooth, M. (2008). Notions of focus anaphoricity. *Acta linguistica Hungarica, 55*, 277–285.

Sahkai, H., Asu, E. L., & Lippus, P. (2022). Prosodic characteristics of canonical and non-canonical questions in Estonian. In *Proceedings of the International Conference on Speech Prosody, Lisbon, Portugal* (pp. 135–139).

Schertz, J., & Clare, E. J. (2019). Phonetic cue weighting in perception and production. *Wiley Interdisciplinary Reviews Cognitive Science, 11*, e1521.

Shen, X. S. (1990). Tonal coarticulation in Mandarin. *Journal of Phonetics, 18*, 281–295.

Shih, C. (1997). Declination in Mandarin. In *Proceedings of the intonation: Theory, models and applications, proceedings of an ESCA workshop* (pp. 293–296). Athens, Greece: European Speech Communication Association.

Shih, C. (2000). A declination model of Mandarin Chinese. In A. Botinis (Ed.), *Intonation. Text, speech and language technology* (Vol. 15). Dordrecht: Springer.

Sorianello, P. (2018). Tra prosodia e pragmatica: Il caso delle domande retoriche. *Studi e Saggi Linguistici LVI, 2*(2), 39–71.

Sorianello, P. (2019). 'A che serve saperlo?' Funzioni pragmatiche e variazioni intonative della domanda retorica. In E. Nuzzo & I. Vedder (Eds.), *Studi AItLA 9: Lingua in contesto. La prospettiva pragmatica* (pp. 89-108). Milan: Officinaventuno.

Turk, A. E., Nakai, S., & Sugahara, M. (2006). Acoustic segment durations in prosodic research: A practical guide. In S. Sudhoff, D. Lenertová, S. Pappert, P. Augurzky, I. Mleinek, N. Richter, ... & J. Schließer (Eds.), *Methods in empirical prosody research*. Berlin, New York: Walter de Gruyter.

van Heuven, V. J., & van Zanten, E. (2005). Speech rate as a secondary prosodic characteristic of polarity questions in three languages. *Speech Communication, 47*, 87–99.

van, J., Rij Hendriks, P., an, H., Rijn Baayen, R. H., Wood & Simon, N. (2019). Analyzing the time course of pupillometric data. *Trends in Hearing, 23*, 1–22.

van Rij, J., Wieling, M., Baayen, R. H., & van Rijn, H. (2017). itsadug: Interpreting time series and autocorrelated data using GAMMs, R package, https://cran.r-project.org/web/packages/itsadug/index.html.

Ward, N. (2019). *The prosodic patterns of English conversation*. Cambridge: Cambridge University Press.

Ward, N., & Gallardo, P. (2017). Non-native differences in prosodic-construction use. *Dialogue and Discourse, 8*, 1.

Wieling, M. (2018). Analyzing dynamic phonetic data using generalized additive mixed modeling: A tutorial focusing on articulatory differences between L1 and L2 speakers of English. *Journal of Phonetics, 70*, 86–116.

Wochner, D. (2022). *Prosody meets pragmatics: A comparison of rhetorical questions, information-seeking questions, exclamatives, and assertions* PhD Thesis. Konstanz: KOPS, University of Konstanz.

Wood, S. N. (2006). *Generalized additive models: An introduction with R*. Boca Raton [u. a.]: CRC Press.

Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 73*, 3–36.

Wood, S. N. (2017). *Generalized additive models: An introduction with R* (2nd ed.). Boca Raton [u.a.]: CRC Press.

Xu, B. (2013). Nandao-questions as a special kind of rhetorical question. *Proceedings of the Sematics and Linguistic Theory (SALT), 22*, 508–526.

Xu, Y. (1997). Contextual tonal variations in Mandarin. *Journal of Phonetics, 25*, 61–83.

Xu, Y. (1999). Effects of tone and focus on the formation and alignment of f(0) contours. *Journal of Phonetics, 27*, 55–105.

Xu, Y. (2013). ProsodyPro - A tool for large-scale systematic prosody analysis. *Proceedings of the Tools and Resources for the Analysis of Speech Prosody (TRASP 2013)*, Aix-en-Provence, France, 7-10.

Xu, Y. (2019). Prosody, tone and intonation. In W. F. Katz & P. F. Assmann (Eds.), *The Routledge Handbook of Phonetics* (pp. 314–356). New York: Routledge.

Xu, Y., & Liu, F. (2006). Tonal alignment, syllable structure and coarticulation: Toward an integrated model. *Italian Journal of Linguistics, 18*, 125–159.

Xu, Y., & Xu, C. (2005). Phonetic realization of focus in English declarative intonation. *Journal of Phonetics, 33*, 159–197.

Yang, Y. (2018). *The two sides of wh-indeterminantes in Mandarin: A prosodic and processing account*. Utrecht: LOT.

Yang, Y., Gryllia, S., & Cheng, L.-L.-S. (2020). Wh-question or wh-declarative? Prosody makes the difference. *Speech Communication, 118*, 21–32.

Yang, Y., Gryllia, S., Pablos, L., & Cheng, L.-L.-S. (2019). Clause type anticipation based on prosody in Mandarin. *International Journal of Chinese Linguistics, 6*, 1–26.
Yuan, J. (2006). Mechanisms of question intonation in Mandarin. In Q. Huo, B. Ma, C. E-S. & H. Li (Eds.), *Chinese Spoken Language Processing* (pp. 19-30): Springer.
Yuan, J., Shen, L., & Chen, F. (2002). The acoustic realization of anger, fear, joy and sadness in Chinese. *Proceedings of the ICSLP-2002*, 2025-2028.
Zahner-Ritter, K., Einfeldt, M., Wochner, D., James, A., Dehé, N., & Braun, B. (2022). Three kinds of rising-falling contours in German wh-questions: Evidence from form and function. *Frontiers in Communication*. https://doi.org/10.3389/fcomm.2022.838955.
Zahner-Ritter, K., Zhao, T., Einfeldt, M., & Braun, B. (2022). How experience with tone in the native language affects the L2 acquisition of pitch accents. *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2022.903879.
Zhang, J., Duanmu, S., & Chen, Y. (2021). Prosodic systems: China and Siberia. In C. Gussenhoven & A. Chen (Eds.), *The Oxford handbook of language prosody*. Oxford, UK: Oxford University Press.