

JOURNAL ARTICLE

Prosody, clause typing, and *wh*-in-situ: Evidence from Mandarin

Stella Gryllia¹, Jenny S. Doetjes¹, Yang Yang² and Lisa L. Cheng^{1,3}

¹ Leiden University Centre for Linguistics, Leiden University, NL

² Center for Linguistics and Applied Linguistics, Guangdong University of Foreign Studies, CN

³ Leiden Institute for Brain and Cognition, Leiden University, NL

Corresponding author: Stella Gryllia (s.gryllia@hum.leidenuniv.nl)

This paper examines the use of prosody for marking upcoming linguistic material in speech production and for anticipating them in speech perception. More specifically, it examines whether in the absence of any overt morphosyntactic cues in the beginning of an utterance, speakers use prosodic means to mark the clause type (declarative or *wh*-question) and whether listeners use these prosodic cues to anticipate the clause type. We report the results of a production and an audio gating experiment. The results of the production experiment show that speakers of Mandarin differentiate declaratives from *wh*-questions right from the onset of the clause by means of duration, F0, and intensity. The results of the audio gating experiment demonstrate that prosody is used by listeners to anticipate the clause type which is intended by the speaker.

Keywords: prosody; *wh*-in-situ; Mandarin

1. Introduction

In everyday activities like doing sports, dancing, or playing music, people seem to be able to predict events or actions of others and respond accordingly. For instance, most of the time, a good goalkeeper is able to foresee the path of the ball and place him/herself in such a way as to keep the ball away from the vicinity of the goal line. The mechanism of anticipation in language is not much different, except that instead of predicting the path of the ball, listeners or readers use a variety of syntactic, morphosyntactic, discourse-semantic, and prosodic cues to predict upcoming linguistic material before it is read or heard (see among others Clark, 2013; Friston, 2010). For instance, in an eye-tracking study that used the Visual World Paradigm, Altmann and Kamide (1999) showed that listeners can predict the complement of a verb on the basis of its selectional properties, while Staub and Clifton (2006) showed in an eye-tracking reading study that readers can predict the upcoming of an ‘*or*-clause’ if they have read an ‘*either*-clause’ before. Similarly, Kamide, Scheepers, and Altmann (2003) offer evidence for anticipation effects based on case marking, Altmann and Kamide (2007) for anticipation of tense, and Arai and Keller (2013) show that verb subcategorization properties result in anticipation of the complement of the verb. In this discussion, we focus on cases of anticipation that are triggered by prosody rather than by morphosyntactic or lexical cues. A large body of literature shows that prosody can be used for anticipation of properties of the upcoming linguistic material in various ways. Ito and Speer (2008), for instance, demonstrate that contrastive accents are used by listeners to anticipate the reference of noun phrases and Brown, Salverda, Dilley, and Tanenhaus (2011) show that lexical segmentation can be

anticipated by prosodic properties of the preceding context. In this paper, we focus on the question of whether prosody can be used in anticipating the clause type of a sentence (declarative or *wh*-question) in the absence of any morphosyntactic cues.

In the literature, a number of production and perception studies focus on the prosodic differences between ordinary declaratives, which are used for making a statement, and declarative questions,¹ which have exactly the same word order but which are used as questions. In English, declarative questions normally have a rising intonation, while an ordinary declarative is characterized by a falling intonation (see for instance Gunlogson, 2002). Recently, it has been shown that ordinary declaratives and declarative questions differ not only with respect to their boundary tone and their nuclear pitch accent, but also with respect to the pre-nuclear area (see for instance Haan, van Heuven, Pacilly, & van Bezooijen, 1997 for Dutch; Petrone & D'Imperio, 2008 for Italian; Petrone & Niebuhr, 2014 for German; Baltazani, Kainada, Lengeris, & Nikolaidis, 2015 for Greek; and Cooper, 2015 for Welsh). Moreover, perception studies confirm that differences between declarative questions and ordinary declaratives can be perceived before hearing the boundary tone of the tune, which in many languages is considered to be the main prosodic cue for a declarative question (see for instance van Heuven & Haan, 2002 for Dutch; Petrone & Niebuhr, 2014 for Northern Standard German; and Face, 2007 for Castilian Spanish). On the other hand, Heeren, Bibyk, Gunlogson, and Tanenhaus (2015) did not find any strong evidence for the use of early prosodic cues by listeners to distinguish ordinary declaratives from declarative questions; the listeners mainly relied on the boundary tone (H% or L%) for differentiating the two types. As indicated by the authors, the pre-nuclear area was very short and the stimuli, which were of the form '*Got a + disyllabic NP*,' possibly in combination with the game set-up of the experiment, resulted in a bias for a question interpretation. This bias and the short length of the pre-nuclear area may have overridden weak cues at the beginning of the sentence.

The aforementioned studies have investigated the use of prosodic cues by speakers and listeners for distinguishing ordinary declarative sentences and questions with a declarative word order. These questions differ from interrogative *yes-no* questions, which are syntactically marked as interrogatives, e.g., in English by means of inversion of the verb and the subject. Gunlogson (2002) argues that declarative questions are a special use of declaratives rather than a special type of interrogative. Contrary to interrogative *yes-no* questions, declarative questions have declarative syntax, and they convey commitment to their propositional content. The commitment to their propositional content introduces the bias that distinguishes declarative questions from neutral interrogative *yes-no* questions. Gunlogson claims that the rise that normally characterizes the intonation of declarative questions attributes this commitment to the addressee rather than to the speaker, while a falling intonation signals that the commitment comes from the speaker. Addressee commitment turns out to depend on strong contextual cues that make this commitment plausible. For instance, a declarative question of the form '*It is raining?*' can only be used in a context in which there is a mutual understanding that the speaker has reasons to believe that it is raining and that the addressee is in a position to confirm this. To conclude, declarative questions impose strong conditions on the context in which they are uttered and the prosodic cues that discriminate ordinary declaratives and declarative questions reflect pragmatic differences rather than differences in syntactic clause type.

¹ In the literature, usually the term *yes-no* questions or unmarked *yes-no* questions is used. Here, we want to emphasize the nature of such questions, and to reflect recent developments in the field by calling these questions declarative questions, following Gunlogson (2002).

In this paper, we compare declaratives with *wh*-in-situ questions. As in the case of declarative questions, *wh*-in-situ questions are at the beginning of the sentence string identical to declarative sentences (if we do not use subject question words or pre-verbal *wh*-adverbials), but contrary to declarative questions, they are interrogatives from a syntactic point of view and they denote questions rather than propositions. Mandarin is a genuine *wh*-in-situ language (Huang, 1982; Cheng, 1991) in which the *wh*-phrase occupies the same linear position as its non-interrogative counterpart, as shown in (1) and (2), where *shénme* ‘what’ and *júzi* ‘oranges’ appear in the post-verbal object position.

- (1) *Bái Wēi zúotiān bō-le shénme gěi Luó Yīng?* (wh-question)
 Bai Wei yesterday peel-PERF what for Luo Ying
 ‘What did Bai Wei peel for Luo Ying yesterday?’
- (2) *Bái Wēi zúotiān bō-le júzi gěi Luó Yīng.* (declarative)
 Bai Wei yesterday peel-PERF orange for Luo Ying
 ‘Bai Wei peeled oranges for Luo Ying.’

As we can see in (1) and (2), a *wh*-question and its declarative counterpart are string identical except for the *wh*-word (*shénme* ‘what’) and its non-*wh* counterpart (*júzi* ‘orange’ in [2]). This means that in Mandarin there are no overt morphosyntactic cues preceding the *wh*-word that can be used to distinguish a *wh*-question from a corresponding declarative. Mandarin *wh*-questions and their declarative counterparts thus offer unique minimal pairs for us to investigate prosodic cues for distinguishing declarative and interrogative sentences.² In particular, from (1) and (2), we investigate the region before the *wh*-word for the prosodic cues. The research questions addressed in this paper are the following:

- (3) Do speakers of Mandarin prosodically mark the clause type (*wh*-question versus declarative) in the pre-*wh*-word contour?
- (4) If so, do listeners use this prosodic marking to anticipate the clause type before reaching the *wh*-word or its non-*wh* counterpart?

The paper is organized as follows. In Section 2, we present the results of a production experiment (Experiment 1) that we conducted to address the question in (3). Section 3 reports the results of an audio gating experiment (Experiment 2) that was run to tackle the question in (4). Section 4 discusses the results and draws conclusions about the prosodic marking of the *wh*- versus declarative clause types in Mandarin and about the use of prosodic cues for clause type anticipation, as well as the interaction between tones and intonation.

² A reviewer suggests that we also compare the *wh*-in situ questions with declarative yes-no questions directly, in order to exclude the possibility that the prosodic properties of *wh*-in situ questions and declarative questions are in certain respects similar and as such could reflect pragmatic similarity between the two sentence types. In such a situation, the observed differences between declarative statements and *wh*-in situ questions might be pragmatic in nature as well. Even though we acknowledge that this is a potential limitation of our study, we show below that the types of prosodic cues that we found for Mandarin *wh*-in situ differ from the ones that are described in the literature for Mandarin declarative questions. Moreover, we suspect that the strong contextual conditions on the use of declarative questions (see above) can lead to confounds when these are compared with *wh*-questions.

2. The phonetics of Mandarin *wh*-questions and declaratives

Previous studies on the intonation of questions in Mandarin have mainly focused on various types of yes-no questions. Crucially, these studies use ‘syntactically unmarked yes-no questions’ (i.e., declarative questions, see above), yes-no questions with the yes-no particle *ma*, and in some cases the semantically neutral A-not-A question is also included.³ Most of the studies report a general higher pitch level in declarative (yes-no) questions than statements (De Francis, 1963; Ho, 1977; Shen, 1990), except Tsao (1967), who claimed that yes-no questions do not differ from the corresponding declaratives in terms of pitch level. However, as Liu (2009) notes, these previous studies did not control for sentence focus and some of them made the claims based on impressionistic inspection of data. Here we briefly review two recent studies, both of which include *wh*-questions as well.

The first one is Lee (2005), who investigated three types of questions (declarative questions, *ma*-questions, and *wh*-questions) in contrast with ordinary declaratives. It is reported that with respect to the declarative questions and *ma*-questions, both the expansion of the pitch range and raising of the overall pitch are manifested towards utterance-final position. Concerning *wh*-questions, the overall pitch starts higher than that of statements. Importantly, Lee reported that in longer utterances, neither declarative questions nor *ma*-questions differ from statements in pitch range in the early portion of the utterances. In the case of *wh*-questions, the *wh*-phrase itself has an expanded pitch range, possibly associated with narrow focus.

Liu’s (2009) study also included the three types of questions reported in Lee (2005), but the focus of the questions is controlled (using initial, medial, final, or neutral focus). Liu, similar to many other scholars in previous studies, constructs her stimuli using only one specific tone for the whole sentence. This choice definitely influences the naturalness of the utterances, as Shen (1990) also reported. She found that in the case of neutral focus, the difference between statements and declarative questions is mainly manifested in the final word, which confirms what is claimed in previous literature. Furthermore, pitch raising by question intonation is greater in yes-no questions (both declarative questions and *ma*-questions) than in *wh*-questions. She suggests that this may be related to a separation of incredulity and interrogation.⁴ Given the fact that both declarative questions and *ma*-questions are biased in the sense of Gunlogson (2002), this might also be a reflection of the specific pragmatic properties of biased questions. Finally, it should be noted that due to the study design of investigating focus, including *initial* focus, some *wh*-questions have initial *wh*-phrases.

In short, the aforementioned studies provide evidence for the claim that declaratives differ from biased yes-no questions in terms of prosody; the pitch level of these yes-no questions is raised in particular towards the end of the utterance. However, due to the biased nature of these yes-no questions, it could be the case that the higher pitch level is associated with the pragmatic interpretation of these questions. As for *wh*-questions, there is some indication that *wh*-questions also have a higher overall pitch, with the *wh*-phrase getting the expanded pitch range. Nonetheless, the *wh*-questions investigated are either followed by the particle *ne* (Lee, 2005),⁵ or they sometimes start with the *wh*-phrase, making it hard to draw conclusions concerning anticipation or recognition

³ Neither syntactically unmarked yes-no questions nor *ma*-questions are neutral yes-no questions. Unmarked yes-no questions are the same as declarative questions discussed in Section 1, and Li and Thompson (1981) already reported that *ma*-questions are non-neutral questions which require certain presuppositions similar to the ones of declarative questions. A-not-A questions are the most neutral yes-no questions in Mandarin; either the verb is in the A-not-A form (i.e., V-not-V) or some adverbs preceding the verb can also be in the A-not-A form.

⁴ Liu (2009) rightly considers declarative questions and *ma*-questions to be biased questions.

⁵ Li (2006) claims that *ne* is an evaluative marker, even in the case of *wh*-questions. In other words, *wh*-questions with *ne* can be considered to be less neutral than *wh*-questions without *ne*.

of *wh*-questions. The production experiment reported in this paper is the first study that zooms in on the prosodic properties of *wh*-questions without *ne* in Mandarin. In particular, the study reported here investigates the prosodic properties of the part of the *wh*-questions preceding the *wh*-phrase, making it possible for us to evaluate for the first time whether or not there are specific prosodic properties of the pre-*wh* region in *wh*-in situ questions that listeners could use to distinguish these questions from declaratives before the *wh*-expression is reached.

2.1. Stimuli

As the central question of our study concerns whether the prosodic properties in the pre-*wh*-word contour in a *wh*-question differ from the prosodic properties of a string identical declarative (except for the *wh*-word slot), we constructed two clause types, namely, *wh*-questions and declaratives. Each stimulus consisted of 11 syllables. Furthermore, we kept the tonal composition of the stimuli constant across items and clause types for all constituents, except for the verb. For the verb (monosyllabic) we included all four tones to ensure that we have more natural stimuli. In total we constructed 56 stimuli (7 exemplars \times 4 verb tones \times 2 clause types). An example set of stimuli is given in (5): (5a) is a *wh*-question, while (5b) is the declarative counterpart, and (5c) gives the general structure of the stimuli and indicates the mapping of the first six syllables and their tones.

- (5) a. *Luó Wēi qiántiān mǎi-le shénme gěi Lú Yīng?*
 Luo Wei the.day.before.yesterday buy-PERF what for Liu Ying
 ‘What did Luo Wei buy for Liu Ying the day before yesterday?’
- b. *Luó Wēi qiántiān mǎi-le tízi gěi Lú Yīng.*
 Luo Wei the.day.before.yesterday buy-PERF grapes for Liu Ying
 ‘Luo Wei bought grapes for Liu Ying the day before yesterday.’
- c. Subject Adverb Verb-PERF Direct Object Indirect Object
 S1S2 S3S4 S5–S6
 T2T1 T2T1 T1/T2/T3/T4-T0

For the subject and the indirect object, we used common Chinese disyllabic proper names. Moreover, for introducing the indirect object we used the verb *gěi* ‘give’ which is a typical element for introducing an indirect object and in this case, it functions like a preposition, and it has a low tone (T3). We also included disyllabic temporal adverbs. In the case of the direct object, the first syllable of the *wh*-word has a rising tone (T2), while the second syllable has a neutral tone (T0), and the same tonal pattern was used for making the corresponding non-question counterpart (e.g., *tízi* ‘grapes’). Target stimuli were intermingled with fillers and a pseudo-randomized list of stimuli was prepared for every participant using Praat (Boersma & Weenink, 2017). A list of the target stimuli is given in Appendix A.

2.2. Procedure

The recordings took place in a sound-proof booth in a lab of the Department of Foreign Languages and Literatures at Tsinghua University in Beijing. Participants were seated in front of a Dell laptop screen at an approximate distance of 50cm, wearing a head-worn vocal microphone (Shure SM10A), and the utterances were recorded with the microphone connected to an external sound card using Audacity software (sampling rate 44.1kHz, 16bit, mono). The stimuli were presented on-screen using Praat, and the presentation pace was controlled by the experimenter. Participants were instructed to first read the sentence on screen silently to understand its meaning and then utter it as if they were engaged into a conversation.

2.3. Participants

Forty participants (23 female and 17 male, \bar{x} age = 21 years old), students at Tsinghua University were paid to participate in the experiment. All were native speakers of Beijing Mandarin, born and raised in Beijing and had normal or corrected-to-normal vision.

2.4. Analysis

2.4.1. Acoustic analysis

A total of 2240 utterances (56 target stimuli \times 40 speakers) was inspected for disfluencies, unnatural pausing, or slips of the tongue, and 171 tokens were excluded from any further analysis. We then analyzed the 2069 utterances with respect to duration, F0, and intensity.

Duration. In every utterance, we manually marked the onset and the offset of the first six syllables in Praat (Boersma & Weenink, 2017), and used a script to extract the duration of each syllable (see **Figure 1**). Note that syllables in Mandarin are words (Duanmu, 2000, 2011). We also calculated the speakers' speech rate adding up the duration of the first six syllables and dividing the outcome by six.

F0. For F0, we obtained a number of tone-specific measurements. Following among others, Duanmu (2000), Chen and Gussenhoven (2008), van de Weijer and Sloos (2014), we represent the four lexical tones in Mandarin (T1, T2, T3, T4) with H(igh) and L(ow); so T1 which is a high level tone, is represented as H, T2, which exhibits a rise, is represented as LH, T3, which is a low tone,⁶ is represented as L, and T4, which exhibits a fall, is represented as HL. We operationalized these phonological representations making the following measurements. For the static tones T1 and T3 (Xu, 1999), following Chen (2010), we measured F0 at the syllable offset (see **Figure 1**), measuring F0-maximum for T1 and F0-minimum for T3. For the dynamic tone T2, we measured the F0-minimum (beginning of the F0-rise) and the F0-maximum (end of the F0 rise). To obtain these measurements, we first identified the F0-maximum of the tone unit which occurred at the syllable offset, we then inspected the F0 leftwards to identify the onset of the F0-rise (F0-minimum). Similarly, for the dynamic tone T4, we measured the F0-maximum (beginning of the fall) and the F0-minimum (end of the fall). After identifying the F0-maximum which occurred at the syllable onset, we inspected the F0 rightwards to identify the offset of the F0-fall (see **Figure 1**). In syllable six (S6) the perfective marker *le* has neutral tone (T0). As the realization of the neutral tone (T0) is influenced by the preceding tone, following Li (2002), we chose our measure points depending on the preceding tone. We measured first the F0-maximum and then the F0-minimum when *le* followed a verb that bore T1, T2, or T4, and we measured first the F0-minimum and then the F0-maximum when *le* followed a verb that bore T3. We also calculated the F0-range in the pre-*wh*- or corresponding non *wh*-word contour subtracting the F0-minimum of T2 in the first syllable of the utterance (S1) from the F0-max of T0 in the sixth syllable of the utterance (S6).

To reduce speaker variation, F0 values in Hz were converted into semitones (ST). For female speakers we used the formula in (6), while for male speakers we used the formula in (7).

$$(6) \quad ST = 12 \log_2(\text{Hz}/100)$$

$$(7) \quad ST = 12 \log_2(\text{Hz}/50)$$

⁶ Tone 3 is a dipping tone when produced in isolation or on the final syllable of an utterance followed by a pause; in a stream of speech Tone 3 is usually produced with a low tone contour (see Cheng, 1968).

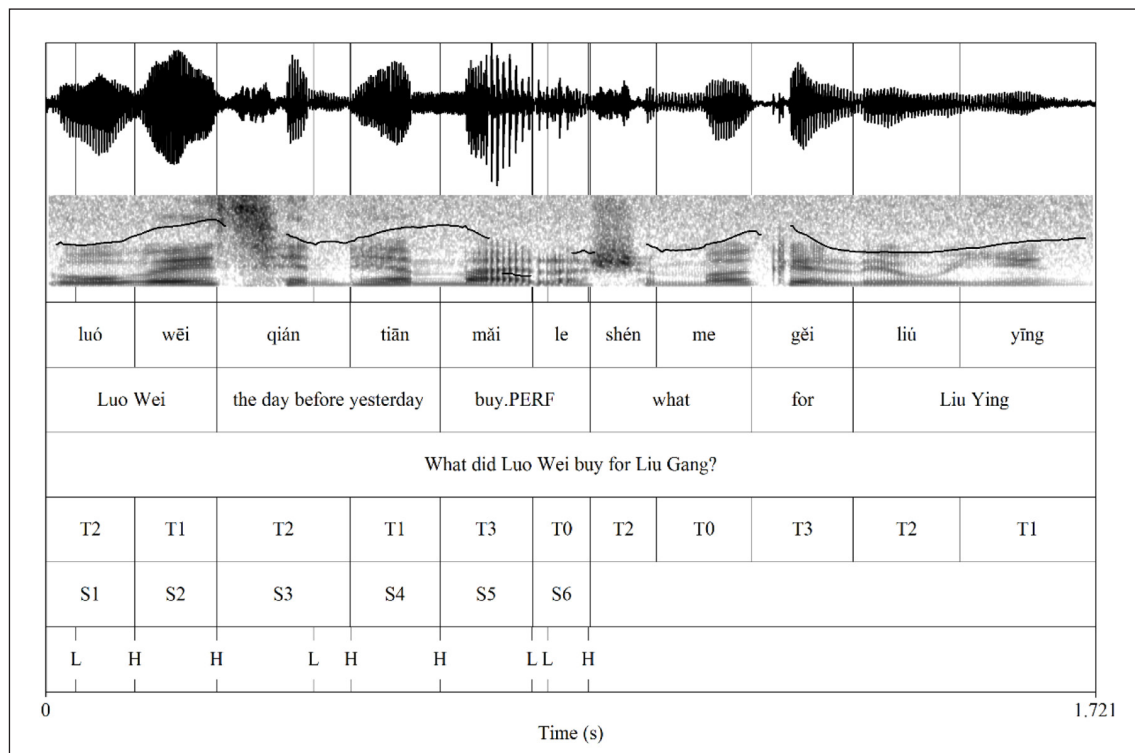


Figure 1: A waveform and a spectrogram with a superimposed F0 curve of a *wh*-question. The tones of the syllables S1–S6 and the corresponding F0 measurement points are indicated in TextTier 4 and PointTier 6.

Intensity. We measured the syllable intensity range, defined as Syllable Maximum Intensity—Syllable Minimum Intensity, rather than the mean syllable intensity, as this measure is more informative and allows us to capture any differences in intensity between declaratives and *wh*-questions (see also Titze, 1988; Chen, 2005; Ouyang & Kaiser, 2013).

2.4.2. Statistical analysis

We ran a series of linear mixed-effect models using the *lmer* function of the *lme4* package (Bates, Mächler, Bolker, & Walker, 2015) in R (R Core Team, 2017).

Duration. We first ran a null model (m0) with *duration* in milliseconds as the dependent variable, and *speaker* and *item* as random factors. Then, we ran a second model (m1) in which we included in addition the *syllable* [S1, S2, S3, S4, S5, S6] as a fixed factor. A third model (m2) included also *type of the clause* [declarative, *wh*-question] as a fixed factor, while the final model (m3) included *duration* as the dependent variable, *type of the clause*, *syllable*, the interaction *type of the clause* × *syllable* as fixed factors, and *by-speaker* and *by-item* random intercepts. Models with maximal random effects failed to converge. All four duration models were compared for model fit; see **Table 1**. In the results section, we present the outcomes of the final model (Duration:m3) that performed best.

Speech rate. We started with a null model (m0) which included *speech rate* as a dependent variable, and *speaker* and *item* as random factors. Then, we ran a second model (m1) in which we included the *type of the clause* [declarative, *wh*-question] as a fixed factor and *by-speaker* and *by-item* random intercepts. A model with maximal random effects failed to converge, and the second model was found to perform best; see **Table 1**.

Table 1: Model fit measures for duration and speech rate models.

| | AIC | BIC | Log-likelihood | Deviance | x² | df | P |
|---------------|------------|------------|-----------------------|-----------------|----------------------|-----------|----------|
| <i>Models</i> | | | | | | | |
| Duration:m0 | 132940 | 132970 | -66466 | 132932 | | | |
| Duration:m1 | 124688 | 124755 | -62335 | 124670 | 8261.824 | 5 | <0.001 |
| Duration:m2 | 124627 | 124701 | -62303 | 124607 | 63.632 | 1 | <0.001 |
| Duration:m3 | 124616 | 124727 | -62293 | 124586 | 20.951 | 5 | <0.001 |
| SpeechRate:m0 | 17344 | 17367 | -8668.2 | 17336 | | | |
| SpeechRate:m1 | 17284 | 17313 | -8637.2 | 17274 | 61.997 | 1 | <0.001 |

F0. We ran two series of analyses, one for F0min measurements and one for F0max measurements. For F0min, similar to duration, we first ran a null model (m0) with *F0min* in semitones as the dependent variable, and speaker and item as random factors. Then, we ran a second model (m1) in which we included in addition the syllable [S1, S3, S5, S6]⁷ as a fixed factor. A third model (m2) included also type of the clause [declarative, *wh*-question] as a fixed factor, while the fourth model (m3) included also the interaction syllable by type of the clause. The final model (m4) included F0min as the dependent variable, type of the clause, syllable, *tone* [T2, T3, T4], and the interaction syllable by type of the clause as fixed factors, and by-speaker and by-item random intercepts. Models with maximal random effects failed to converge. All five F0min models were subsequently compared for model fit; see **Table 2**. In the results section, we present the outcomes of the final model (F0min:m4) that performed best. A similar procedure was followed for F0max. In **Table 2** we report the model fit for all five F0max models. In the results section, we present the outcomes of the final model (F0max:m4) that performed best.

F0-range. We started with a null model which included *F0-range* as a dependent variable, and speaker and item as random factors. Then, we ran a second model in which we

Table 2: Model fit measures for F0min, F0max, and F0-range models.

| | AIC | BIC | Log-likelihood | Deviance | x² | df | P |
|---------------|------------|------------|-----------------------|-----------------|----------------------|-----------|----------|
| <i>Models</i> | | | | | | | |
| F0min:m0 | 34043 | 34070 | -17017 | 34035 | | | |
| F0min:m1 | 33228 | 33276 | -16607 | 33214 | 821.33 | 3 | <0.001 |
| F0min:m2 | 33127 | 33182 | -16556 | 33111 | 102.32 | 1 | <0.001 |
| F0min:m3 | 33109 | 33185 | -16544 | 33087 | 23.97 | 3 | <0.001 |
| F0min:m4 | 32921 | 33010 | -16448 | 32895 | 192.04 | 2 | <0.001 |
| F0max:m0 | 52965 | 52994 | -26478 | 52957 | | | |
| F0max:m1 | 52193 | 52258 | -26087 | 52175 | 781.81 | 5 | <0.001 |
| F0max:m2 | 52153 | 52226 | -26067 | 52133 | 41.859 | 1 | <0.001 |
| F0max:m3 | 52124 | 52234 | -26047 | 52094 | 38.561 | 5 | <0.001 |
| F0max:m4 | 51146 | 51270 | -25556 | 51112 | 982.13 | 2 | <0.001 |
| F0Range:m0 | 11176 | 11198 | -5584.0 | 11168 | | | |
| F0Range:m1 | 11173 | 11200 | -5581.3 | 11163 | 5.2747 | 1 | <0.05 |
| F0Range:m2 | 11120 | 11165 | -5552.1 | 11104 | 58.568 | 3 | <0.001 |
| F0Range:m3 | 11120 | 11165 | -5552.1 | 11104 | 58.568 | 3 | <0.001 |

⁷ Syllable two (S2) and syllable four (S4) are not contributing any data, as S2 and S4 bear T1 (High level tone).

included in addition the type of the clause [declarative, *wh*-question] as a fixed factor. A third model included also the *tone of the verb* [T1, T2, T3, T4] as a fixed factor, while the final model included F0-range as a dependent variable, type of the clause, tone of the verb and their *interaction* as fixed factors, and by-speaker and by-item random intercepts. In the results section we present the results of the last model which was found to perform best, see **Table 2**.

Intensity. We first ran a null model with *intensity range* in dB as the dependent variable, and speaker and item as random factors. Then, we ran a second model in which we included in addition the syllable [S1, S2, S3, S4, S5Tone1, S5Tone2, S5Tone3, S5Tone4, S6Tone1, S6Tone2, S6Tone3, S6Tone4] as a fixed factor. A third model included also type of the clause [declarative, *wh*-question] as a fixed factor, while the final model included intensity range as the dependent variable, type of the clause, syllable and their interaction as fixed factors, and by-speaker and by-item random intercepts. Models with maximal random effects failed to converge. All four intensity models were compared for model fit; see **Table 3**. In the results section, we present the outcomes of the final model (Intensity:m3), despite the fact that it is not the best fit model.

Table 3: Model fit measures for intensity range models.

| | AIC | BIC | Log-likelihood | Deviance | χ^2 | df | <i>p</i> |
|---------------|-------|-------|----------------|----------|----------|----|----------|
| <i>Models</i> | | | | | | | |
| Intensity:m0 | 84645 | 84675 | -42319 | 84637 | | | |
| Intensity:m1 | 74714 | 74825 | -37342 | 74684 | 9953.3 | 11 | <0.001 |
| Intensity:m2 | 74716 | 74835 | -37342 | 74684 | 0.2804 | 1 | 0.5964 |
| Intensity:m3 | 74715 | 74915 | -37330 | 74661 | 22.854 | 11 | 0.018 |

2.5. Results

Duration. **Figure 2** displays the mean duration (in ms) of the first six syllables in *wh*-questions paired to their counterparts in declaratives. As shown in **Figure 2**, the mean syllable duration of declaratives is longer than the mean syllable duration of *wh*-questions.

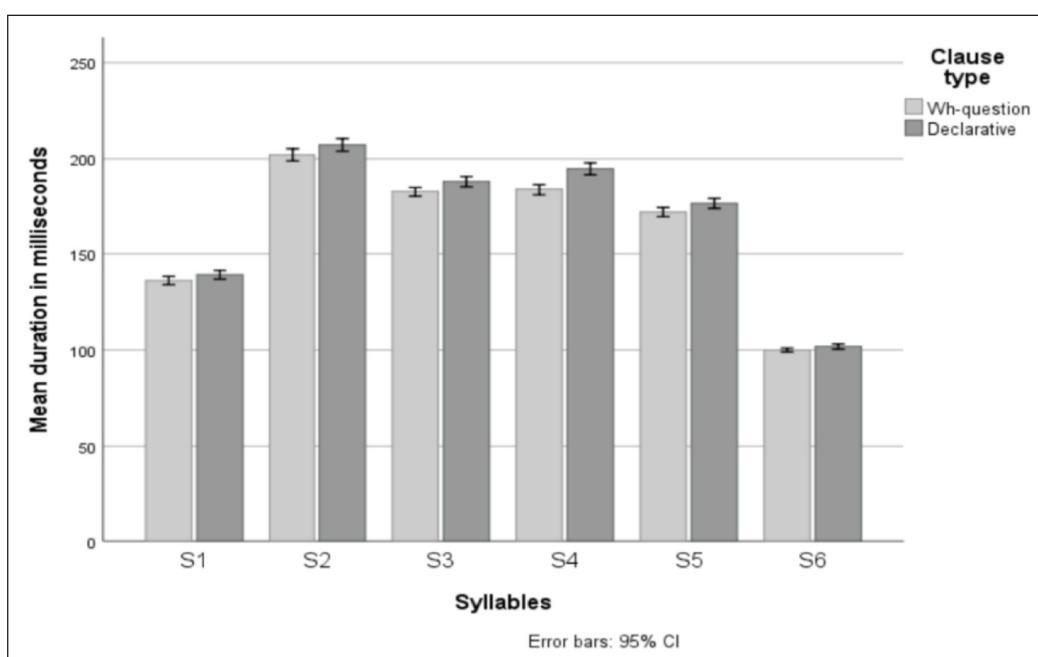


Figure 2: Mean duration of syllables S1-S6 in *wh*-questions and declaratives.

Our results show a significant effect of the type of the clause on syllable duration for syllables S2–S5 [S2: Estimates = 5.15, $SE = 1.594$, $t\text{-value} = 3.233$, $p < 0.002$, S3: Estimates = 5.61, $SE = 1.594$, $t\text{-value} = 3.520$, $p < 0.001$, S4: Estimates = 11.19, $SE = 1.594$, $t\text{-value} = 7.021$, $p < 0.001$, S5: Estimates = 4.49, $SE = 1.594$, $t\text{-value} = 2.814$, $p < 0.006$]. Syllables S1 and S6 did not differ significantly between *wh*-questions and declaratives [S1: Estimates = 2.98, $SE = 1.594$, $t\text{-value} = 1.870$, $p > 0.05$, S6: Estimates = 1.78, $SE = 1.594$, $t\text{-value} = 1.114$, $p > 0.05$]; see Tables 1–6 in Appendix B for a detailed overview.

The type of the clause also had a significant effect on speech rate; the first six syllables of *wh*-questions were uttered faster than the corresponding declarative syllables; see **Figure 3** and Table 7 in Appendix B [Estimates = 5.199, $SE = 0.6554$, $t\text{-value} = 7.933$, $p < 0.001$].

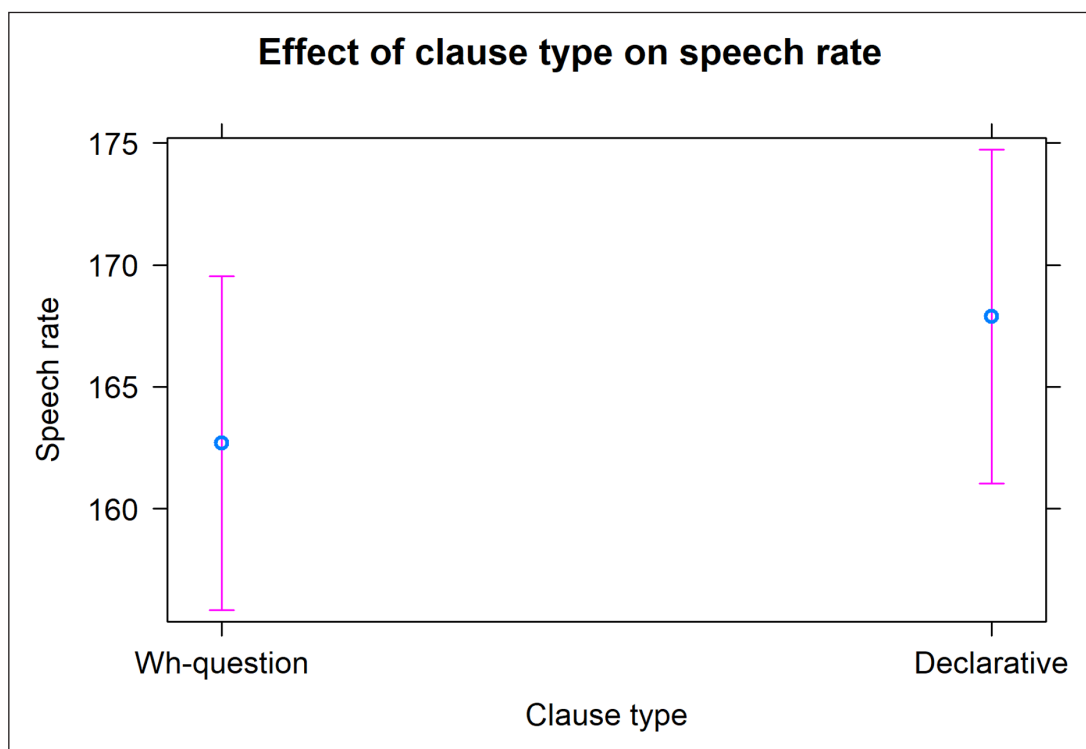


Figure 3: Mean speech rate of syllables S1–S6 for *wh*-questions and declaratives.

F0. **Figure 4** presents the mean values of the *F0* measurement points for syllables S1–S6 in *wh*-questions as compared to their counterparts in declaratives, broken down per verb tone. As shown in **Figure 4**, the mean *F0* values of the *F0* measurement points are higher in *wh*-questions than the corresponding *F0* points in declaratives; the only exception is the H (*F0*-maximum) of the perfective marker *le* when the verb bears T1 or T4.

Our results showed a significant effect of clause type for *F0*min and *F0*max measurements. Specifically for *F0*min, the L of the first syllable of the adverb (S3.T2.*F0*min in **Figure 4**) was significantly lower in declaratives than in the corresponding *wh*-questions (see Table 9 in Appendix B). Moreover, the L of the verb, when the verb bore T2, T3, or T4 (S5.T2.*F0*min, S5.T3.*F0*min, S5.T4.*F0*min in **Figure 4**) was significantly lower in declaratives than in the corresponding *wh*-questions (see Tables 10–12 in Appendix B). The differences between declaratives and *wh*-questions were also visible on the perfective marker *le*; the L was significantly lower in declaratives (see Tables 13–14 in Appendix B).

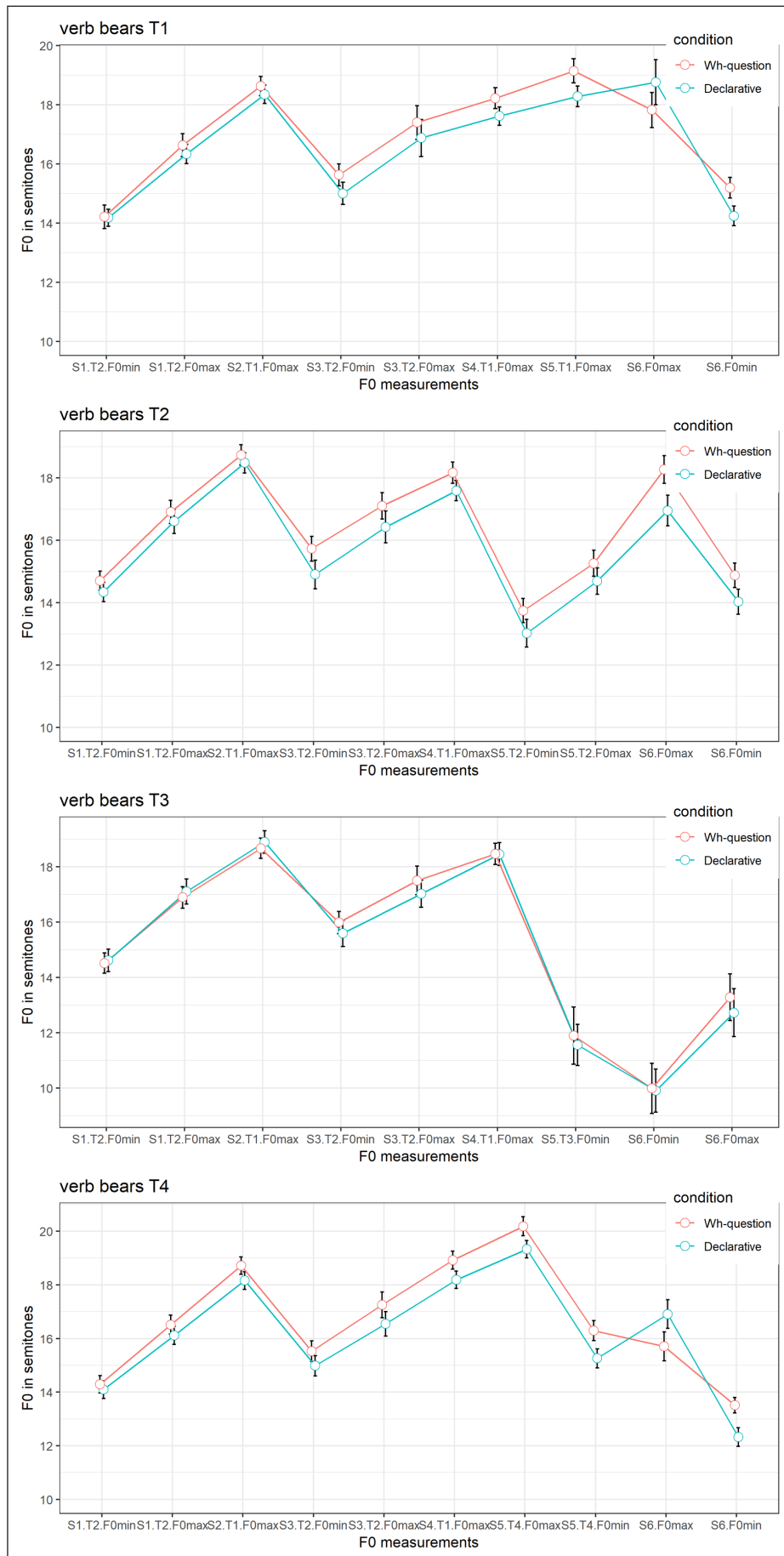


Figure 4: Mean F0 in semitones for syllables S1–S6 in wh-questions and declaratives for all four verb tones (T1–T4). Error bars 95% CI.

For F0max, the H of the second syllable (S2.T1.F0max in **Figure 4**) was significantly lower in declaratives than in the corresponding *wh*-questions (see Table 16 in Appendix B). Moreover, the H of the first and the second syllable of the adverb (S3.T2.F0max, S4.T1.F0max in **Figure 4**) was significantly lower in declaratives than in the corresponding *wh*-questions (see Table 17–18 in Appendix B). Additionally, the H of the verb, when the verb bore T1, T2, or T4 (S5.T1.F0max, S5.T2.F0max, S5.T4.F0max in **Figure 4**) was significantly lower in declaratives than in the corresponding *wh*-questions (see Tables 19–21 in Appendix B).

We also found a significant effect of clause type on F0-range; when the verb bore T1 or T4, declaratives showed a larger F0-range than *wh*-questions. The difference between the two clause types was not significant when the verb was T2. The opposite pattern was observed when the verb bore T3; see **Figure 5** and Tables 24–27 in Appendix B.

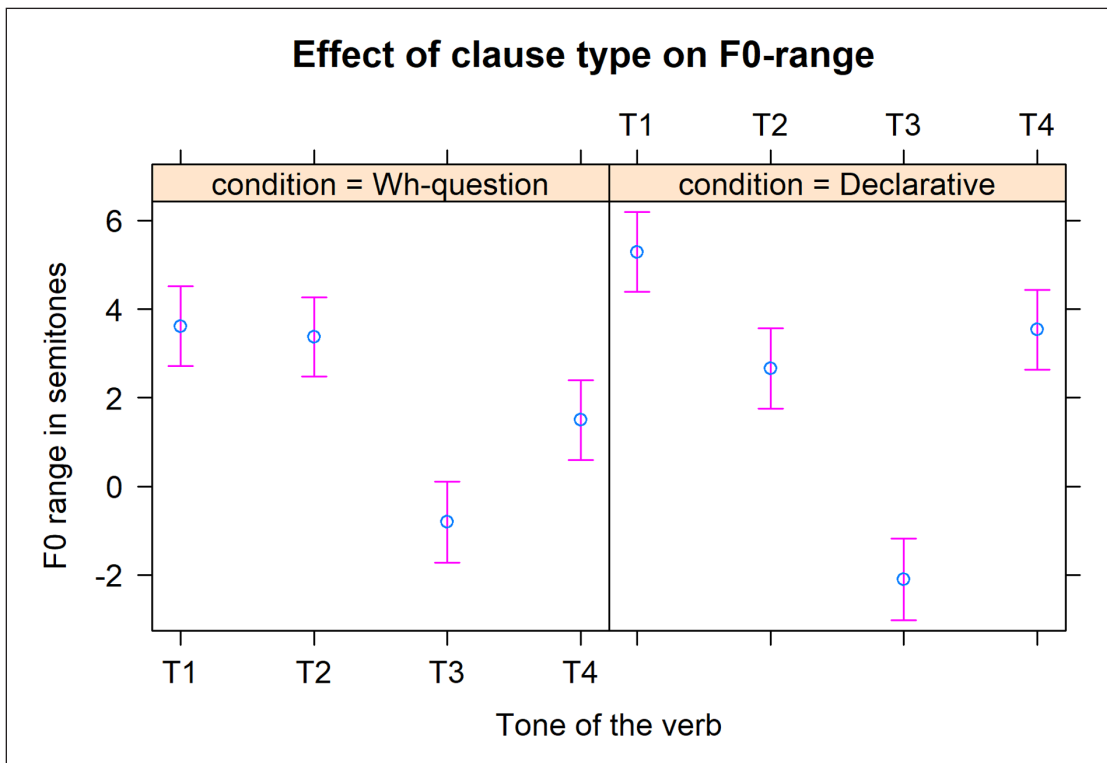


Figure 5: Mean F0-range in *wh*-questions and declaratives for all four verb tones (T1–T4). Error bars 95% CI.

Intensity. Our results show that the mean intensity range at S4 was significantly higher in declaratives than in the corresponding *wh*-questions [Estimates = 0.5345, SE = 0.2133, $t = 2.506$, $p < 0.05$]; see Table 31 in Appendix B. When examining the mean intensity range at S1, S2, and S3, there was no significant difference between the two clause types; see **Figure 6** and Tables 28–30 in Appendix B. When looking at S5, the two clause types do not differ with respect to intensity range. However, when looking at S6, there is a significant difference between *wh*-questions and declaratives, the former having a higher intensity range, when the verb carries T4 [Estimates = -1.0576, SE = 0.4267, $t = -2.478$, $p < 0.05$]; the difference is not significant when the verb bears T1, T2, and T3; see **Figure 7** and Tables 31–39 in Appendix B.

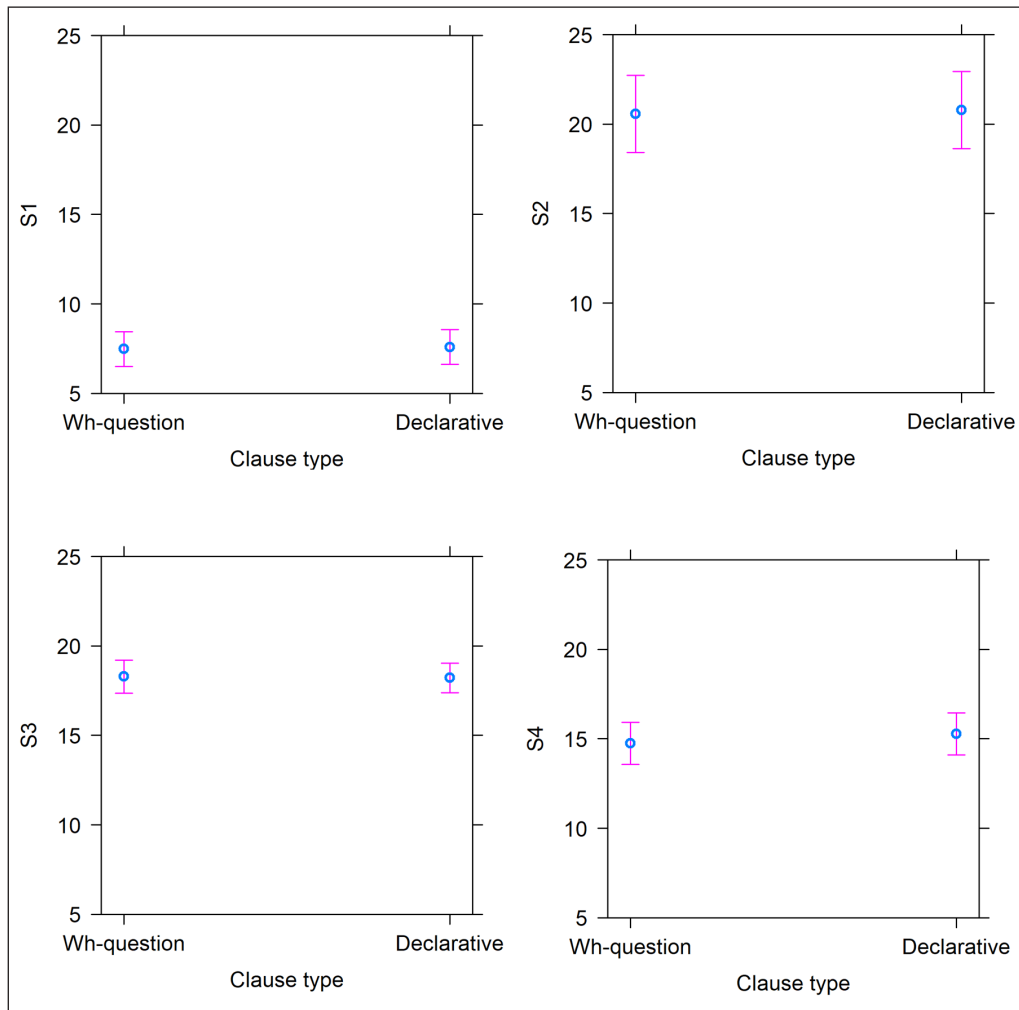


Figure 6: Mean intensity range in dB of syllables S1–S4 in wh-questions and declaratives.

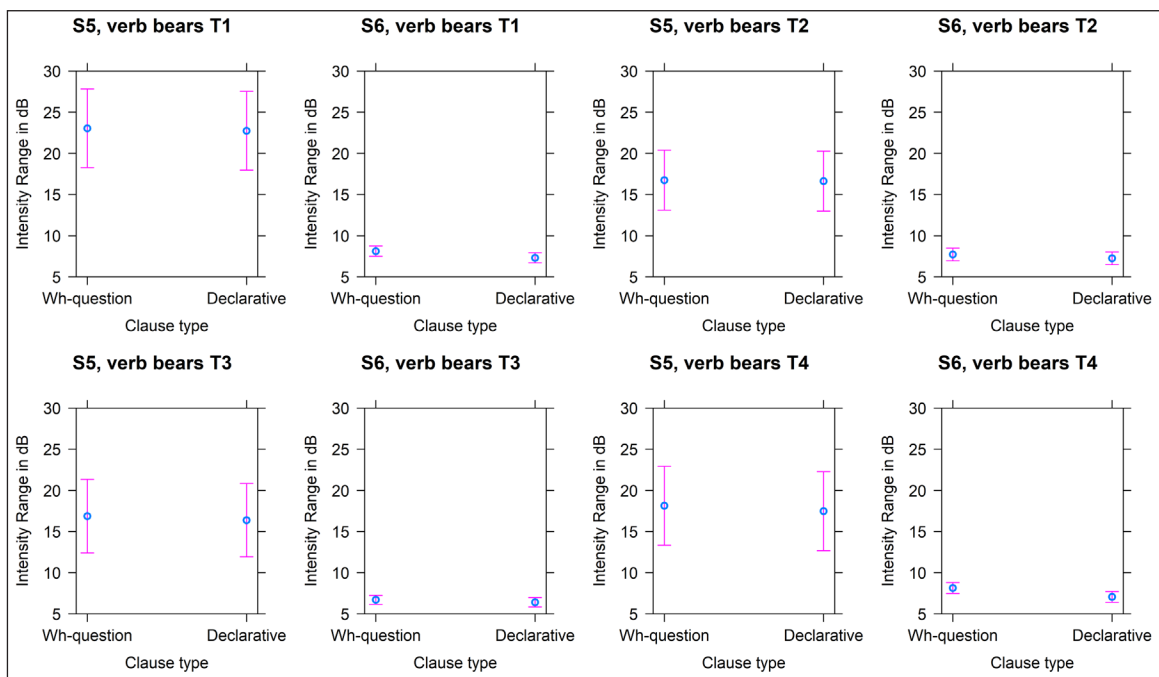


Figure 7: Mean intensity range in dB of syllables S5 and S6 in in wh-questions and declaratives.

2.6. Interim conclusion

The results of the production experiment show that speakers mark prosodically the intended clause type already from the onset of the clause. The prosodic differences between the two clause types are summarized below.

Duration. Looking at the region before the direct object (*wh*-word or its non-*wh*-counterpart) syllables S2–S5 in declaratives are significantly longer than the corresponding syllables in *wh*-questions. In this region, *wh*-questions are uttered significantly faster than their declarative counterparts.

F0. *Wh*-questions have higher F0 than the corresponding declaratives. *Wh*-questions also have a smaller F0 range in the pre-*wh*-word contour in comparison to the corresponding declaratives, when the verb bears T1 or T4.

Intensity. In *wh*-questions, S4 (second syllable of the adverb) has a significantly higher intensity range than the corresponding declaratives.

On the basis of these results, a naturally emerging question is whether listeners use these prosodic cues (duration, F0, and intensity) to anticipate the type of the clause before reaching the *wh*-word or *noun*. To address this question, we conducted an audio perception experiment, which is the focus of the following section.

3. Audio perception experiment: Clause type and anticipation

The findings of the production study show that Mandarin *wh*-questions and declaratives differ prosodically in the region before the direct object (the direct object is a *wh*-word or a non-*wh*-counterpart) and thus suggest that information about the type of the clause is encoded at least prosodically prior to the *wh*-word. A relevant question is whether listeners use the acoustic cues produced by the speakers, and how early in time they interpret the two audio fragments in a distinct way. The audio perception experiment reported here aims at tackling this issue. For setting up this experiment we used a modification of the gating paradigm (Grosjean, 1980; Lahiri & Marlsen-Wilson, 1991). The gating paradigm was initially used for spoken word recognition. The experimenter presented a word repeatedly varying its length and pitch, and participants were asked to write down the word they had heard after each hearing, and indicate their degree of confidence. In the classic gating paradigm, word and prosodic phrase boundaries were manipulated. In later studies, modifications of the gating paradigm have been used for studying the contribution of prenuclear accents to the meaning of the utterance; in these studies, the word and prosodic phrase boundaries stay intact (see Petrone & Niebuhr, 2014).

In our audio perception experiment, participants heard audio fragments and were asked to complete the audio fragment selecting a declarative or question continuation that appeared written on their screen. If listeners use early acoustic cues to anticipate the clause type, then this finding offers some evidence for the use of prosody for the anticipation of the clause type.

3.1. Method

3.1.1. Participants

A total of 36 participants (20 male, 16 female, \bar{x} age = 19 years old) were reimbursed to take part in the experiment. All of them were native speakers of Beijing Mandarin coming from the Beijing area, and none of them reported any hearing disorders.

3.1.2. Stimuli

The third author of the paper who is a native speaker of Beijing Mandarin inspected the production data and evaluated the speakers for naturalness and clear articulation.

After this inspection, we selected a female speaker (20 years old). This speaker was considered to be one of the best speakers of the 40 speakers who participated in the production study reported in Section 2 in the sense that her speech is clear, natural, and with a good pace. Subsequently, the third author inspected the total production data of this female speaker and selected 40 stimuli on the basis of their naturalness. These stimuli consisted of 20 sets; each set included a declarative and its corresponding *wh*-question (20 sets \times 2 clause types). As discussed in Section 2, tones were kept constant across clause types and items for all constituents except for the verb; the tone of the verb varied and all four lexical tones were included. The stimuli were used as basis for constructing the audio fragments. Specifically, each of the 40 stimuli was cut at three different points: at the offset of the subject of the sentence (gate-*a*), at the offset of the adverb (gate-*b*), and at the offset of the perfective marker *le* (gate-*c*)⁸. This resulted in a total of 120 audio fragments. An example of the three gates is given in **Figure 8** and examples (8)–(10).

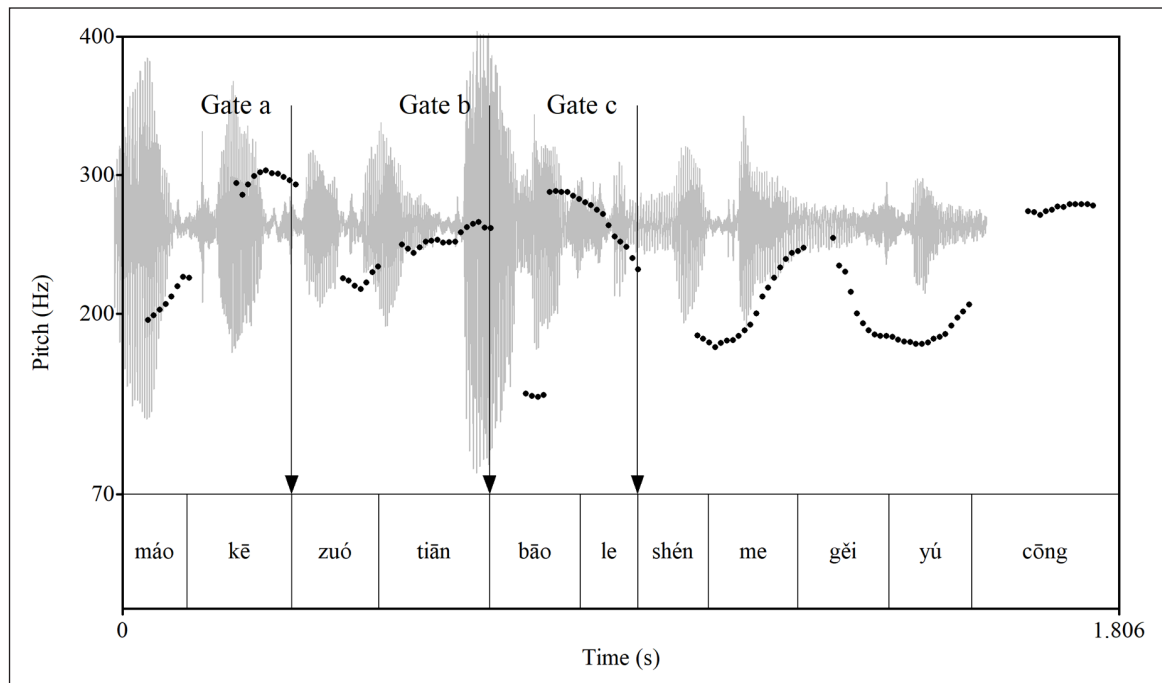


Figure 8: Waveform and F0 contour of the *wh*-in-situ question *Máo Kē zuótiān bāo-le gěi Yú Cōng?* ‘What did Mao Ke peel for Yu Cong yesterday?’

- (8) Gate-a
Máo Kē
 Mao Ke
 ‘Mao Ke’ (proper name)
- (9) Gate-b
Máo Kē zuótiān
 Mao Ke yesterday
 ‘Mao Ke yesterday’

⁸ For reasons of naturalness we decided to respect the word boundaries and therefore cut the stimuli at the offset of words; see Petrone and Niebuhr (2014) for a similar reasoning.

- (10) Gate-c
Máo Kē zuótiān bāo-le
 Mao Ke yesterday peel-PERF
 ‘Mao Ke yesterday peeled’

The number of syllables and the corresponding tones of the three audio fragments is given in (11)–(13).

- (11) Gate-a
 Subject
 S1 S2
 [T2 T1]

- (12) Gate-b
 Subject Adverb
 S1 S2 S3 S4
 [T2 T1 T2 T1]

- (13) Gate-c
 Subject Adverb Verb PERF
 S1 S2 S3 S4 S5 S6
 [T2 T1 T2 T1 T1,T2,T3,T4 T0]

Acoustic properties of the stimuli. As indicated above, the duration and the F0 of the stimuli uttered by this selected speaker are not the mean duration and the mean F0 of the stimuli that were uttered by the 40 speakers. We present here the acoustic properties of the stimuli that were used in gate-c (syllables S1–S6), as these overlap with the stimuli used at the two other gates. Specifically, we present information about the duration, speech rate, F0, and intensity range of the stimuli that were used in the audio perception experiment and briefly indicate the main differences with the results of the production study.

Statistical analysis of the acoustic properties of the stimuli. We ran a series of linear mixed-effects models using the *lmer* function of the *lme4* package (Bates et al., 2015) in R (R Core Team, 2017). Specifically, for every measurement, we first ran a null model with the relevant measurement in milliseconds as the dependent variable and items as random factors. A second model included in addition clause type as a fixed factor. Model fit was compared using the fit measures Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) (Agresti, 2002).

Duration. The total duration (S1–S6) of the fragmented declarative stimuli (\bar{x} = 1003ms) was on average significantly longer than the total duration of the fragmented *wh*-question stimuli (\bar{x} = 952ms), [Estimates = 0.051, SE = 0.013, t = 3.85]. Moreover, the duration of each syllable of the fragmented declarative stimuli was longer than the duration of each syllable of the fragmented *wh*-question stimuli; see **Figure 9**. The difference between the two clause types was statistically significant for S1–S4, as shown in **Table 4**. There was also a difference in S5 and S6, but contrary to what was found in the production study, this difference was not significant.

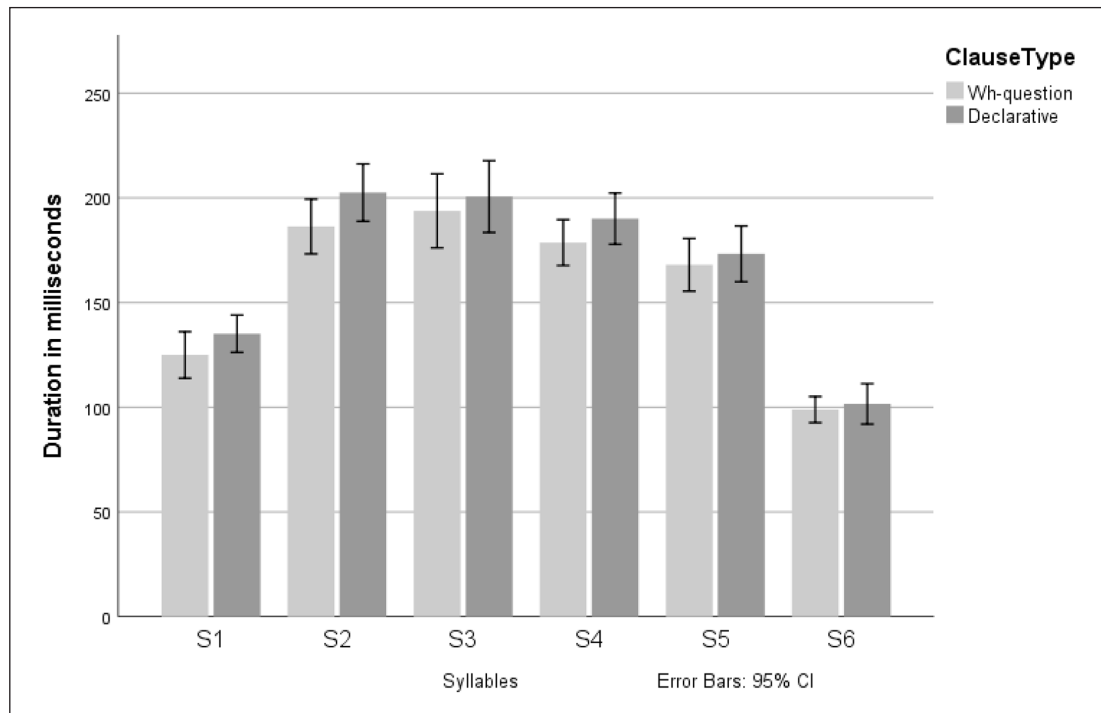


Figure 9: Mean duration of syllables S1-S6 for wh-questions and declaratives.

Table 4: Results of linear mixed effect models on duration.

| | S1 | S2 | S3 | S4 | S5 | S6 |
|-------------------|------------------|---------------|---------------|---------------|---------------|--------------|
| Predictors | Estimates | | | | | |
| (Intercept) | 125.50 | 187.70 | 192.35 | 180.50 | 167.10 | 99.15 |
| | (4.669) | (6.305) | (8.147) | (5.566) | (6.063) | (3.827) |
| | [26.882] | [29.770] | [23.610] | [32.430] | [27.561] | [25.910] |
| | {<0.001} | {<0.001} | {<0.001} | {<0.001} | {<0.001} | {<0.001} |
| Declarative | 9.60 | 14.80 | 8.25 | 9.55 | 6.15 | 2.45 |
| | (4.024) | (5.474) | (3.945) | (4.396) | (5.309) | (4.279) |
| | [2.386] | [2.704] | [2.091] | [2.173] | [1.158] | [0.573] |
| | {0.017} | {0.007} | {0.037} | {0.030} | {0.247} | {0.567} |

Note: Standard error is given in parenthesis (), *t*-value is given in [], and *p*-value is given in { }. Bold indicates significance <0.05.

The fragmented declarative stimuli were also slower in speech rate than the fragmented wh-question stimuli [Estimates = 0.008, *SE* = 0.002, *t*-value = 3.85]; see **Figure 10**.

F0. In general, the *F0* points of the fragmented wh-questions stimuli were higher than the *F0* points of the corresponding declarative stimuli; see **Figure 11**.

This difference between wh-questions and declaratives was significant at the H (*F0*max) of syllable two (S2) which bore T1 and the H (*F0*max) of syllable four (S4) which also bore T1; see **Table 5**. Moreover, there was a significant difference at the H of the verb when the verb bore T1 or T4. When the verb bore T2 or T3, the significant difference was at the first *F0* point of the perfective marker *le*; see **Table 5**. As for the production experiment (see

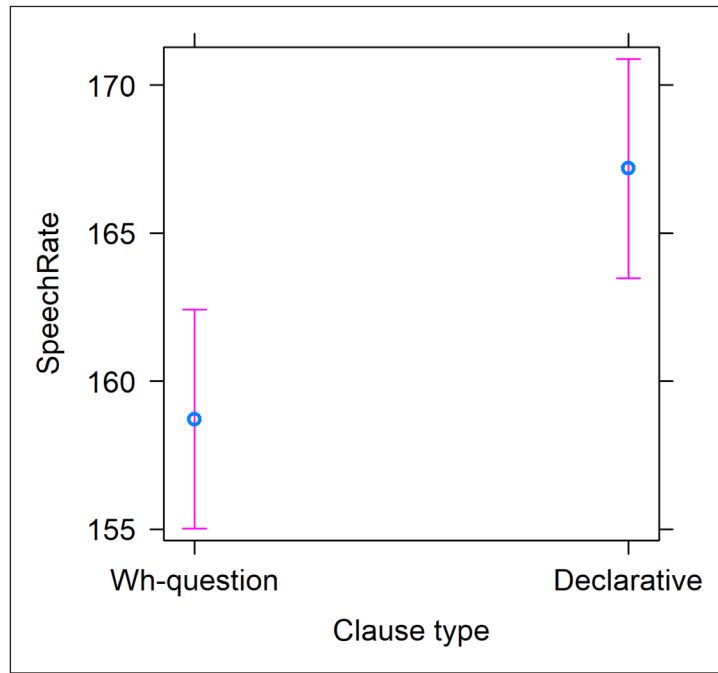


Figure 10: Mean speech rate in wh-questions and declaratives.

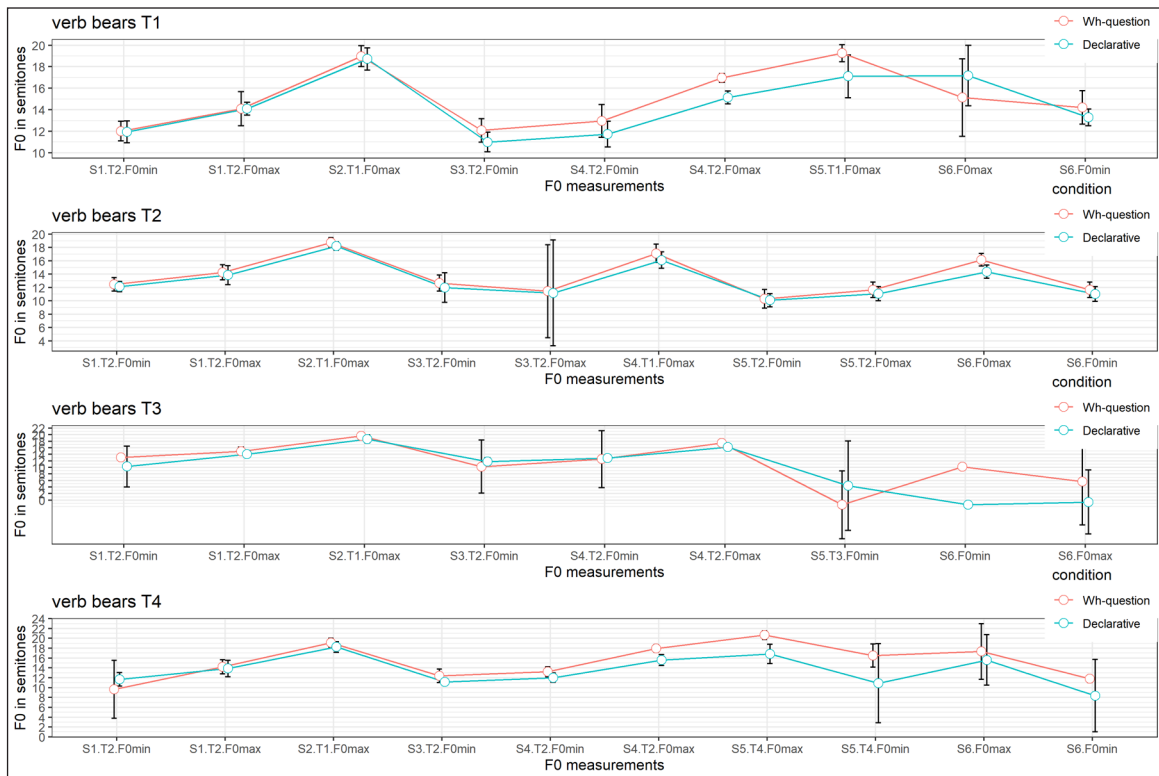


Figure 11: Mean F0 points of fragmented wh-questions and declaratives. Error bars 95%CI.

Figure 4), the F0 points were significantly higher in wh-questions for syllables (S2–S4). When looking at the F0 measurements for S5 and S6 in the production stimuli, we see that the differences between wh-questions and declaratives are significant at the verb (S5); there was no significant difference on the H when the verb bore T2 (S5.T2.F0max). The comparison with the data from the production study shows that in general, the patterns were similar, even though the differences were less often significant in the data used for the perception experiment.

Table 5: Results of linear mixed effect models on F0.

| <i>Syllables</i> | S1 | | S2 | S3 | | S4 |
|-------------------|--|--|---|--|--|--|
| <i>Tones</i> | T2 | | T1 | T2 | T1 | |
| <i>F0-</i> | min | max | max | min | max | max |
| <i>Predictors</i> | <i>Estimates</i> | | | | | |
| (Intercept) | 11.79 (0.67) [17.59] {<0.001} | 14.38 (0.51) [28.169] {<0.001} | 19.70 (0.62) [31.748] {<0.001} | 11.25 (0.78) [14.49] {<0.001} | 12.54 (0.82) [15.23] {<0.001} | 17.38 (0.17) [100.08] {<0.001} |
| Declarative | -0.86 (0.94) [-0.91] {0.363} | 0.29 (0.72) [0.408] {0.683} | -0.70 (0.21) [-3.368] {0.001} | -0.48 (1.10) [-0.44] {0.660} | -0.58 (1.17) [-0.497] {0.619} | -1.62 (0.17) [-9.696] {<0.001} |
| <i>Syllables</i> | S5 | | S6 | S5 | | S6 |
| <i>Tones</i> | T1 | | T2 | T2 | | T2 |
| <i>F0-</i> | max | max | min | min | max | max |
| <i>Predictors</i> | <i>Estimates</i> | | | | | |
| (Intercept) | 17.10 (0.54) [31.41] {0.001} | 17.17 (1.61) [14.79] {0.001} | 13.28 (0.44) [30.024] {<0.001} | 10.09 (0.44) [23.042] {<0.001} | 11.08 (0.40) [27.583] {<0.001} | 14.37 (0.35) [41.438] {<0.001} |
| Declarative | -2.14 (0.77) [2.785] {0.005} | 2.05 (1.03) [1.99] {0.050} | -0.94 (0.62) [1.503] {0.133} | 0.21 (0.52) [0.397] {0.692} | 0.56 (0.44) [1.263] {0.207} | 1.78 (0.41) [4.358] {<0.001} |
| <i>Syllables</i> | S6 | | S5 | S6 | | S5 |
| <i>Tones</i> | T2 | | T3 | T4 | | T4 |
| <i>F0-</i> | min | min | min | max | max | min |
| <i>Predictors</i> | <i>Estimates</i> | | | | | |
| (Intercept) | 11.02 (0.41) [26.691] {<0.001} | 4.39 (4.37) [1.005] {0.315} | 10.18 (0.25) [-4.94] {<0.001} | -0.67 (4.2) [-0.15] {0.885} | 16.84 (0.54) [31.036] {0.001} | 10.86 (2.13) [5.091] {<0.001} |
| Declarative | 0.62 (0.48) [1.297] {0.195} | 5.81 (3.58) [1.63] {0.104} | -11.40 (0.33) [34.597] {<0.001} | -6.28 (2.93) [2.14] {0.032} | -3.82 (0.76) [4.977] {0.001} | -5.64 (2.71) [2.086] {0.037} |
| <i>Syllables</i> | S6 | | | | | |
| <i>Tones</i> | T4 | | | | | |
| <i>F0-</i> | max | min | | | | |
| <i>Predictors</i> | <i>Estimates</i> | | | | | |
| (Intercept) | 15.61 (1.94) [8.033] {<0.001} | 11.78 (1.88) [4.459] {<0.001} | | | | |
| Declarative | -1.72 (2.75) [-0.626] {0.531} | -3.42 (2.49) [-1.373] {0.170} | | | | |

Note: Standard error is given in parenthesis (), t-value is given in [], and p-value is given in { }. Bold indicates significance <0.05.

Intensity. The mean intensity range of syllables S1–S6 was not significantly different between declaratives and *wh*-questions, with the exception of S5 when the verb bore T2; see **Table 6** and **Figure 12**. In the data of the production experiment, we only found a difference in intensity range for S6.

Table 6: Results of linear mixed effect models on intensity range.

| | S1 | S2 | S3 | S4 | S5 | S6 |
|-------------------|---|---|---|---|---|--|
| Tones | T1 | | | | | |
| Predictors | Estimates | | | | | |
| (Intercept) | 7.13 (0.96) [7.40] {<0.001} | 22.65 (1.46) [15.48] {<0.001} | 18.82 (0.74) [25.29] {<0.001} | 16.09 (0.72) [22.28] {<0.001} | 24.40 (4.80) [5.085] {<0.001} | 7.52 (0.95) [7.867] {<0.001} |
| Declarative | 1.10 (0.55) [1.98] {0.048} | 1.37 (0.82) [1.675] {0.094} | 1.19 (0.93) [1.28] {0.201} | 0.06 (0.75) [0.08] {0.933} | -0.15 (1.85) [-0.080] {0.936} | 0.14 (0.81) [0.175] {0.861} |
| | S5 | S6 | S5 | S6 | S5 | S6 |
| Tones | T2 | | T3 | | T4 | |
| Predictors | Estimates | | | | | |
| (Intercept) | 16.90 (3.30) [5.128] {<0.001} | 7.89 (1.08) [7.323] {0.001} | 15.58 (3.55) [4.395] {0.001} | 6.72 (0.95) [7.109] {<0.001} | 21.97 (3.61) [6.095] {<0.001} | 9.22 (1.51) [6.106] {<0.001} |
| Declarative | 2.75 (0.71) [3.889] {<0.001} | -0.96 (1.27) [-0.751] {0.452} | 0.13 (0.77) [0.165] {0.869} | -0.44 (1.38) [-0.330] {0.742} | -1.43 (1.79) [-0.800] {0.424} | -2.80 (1.59) [-1.758] {0.079} |

Note: Standard error is given in parenthesis (), *t*-value is given in [], and *p*-value is given in { }. Bold indicates significance <0.05.

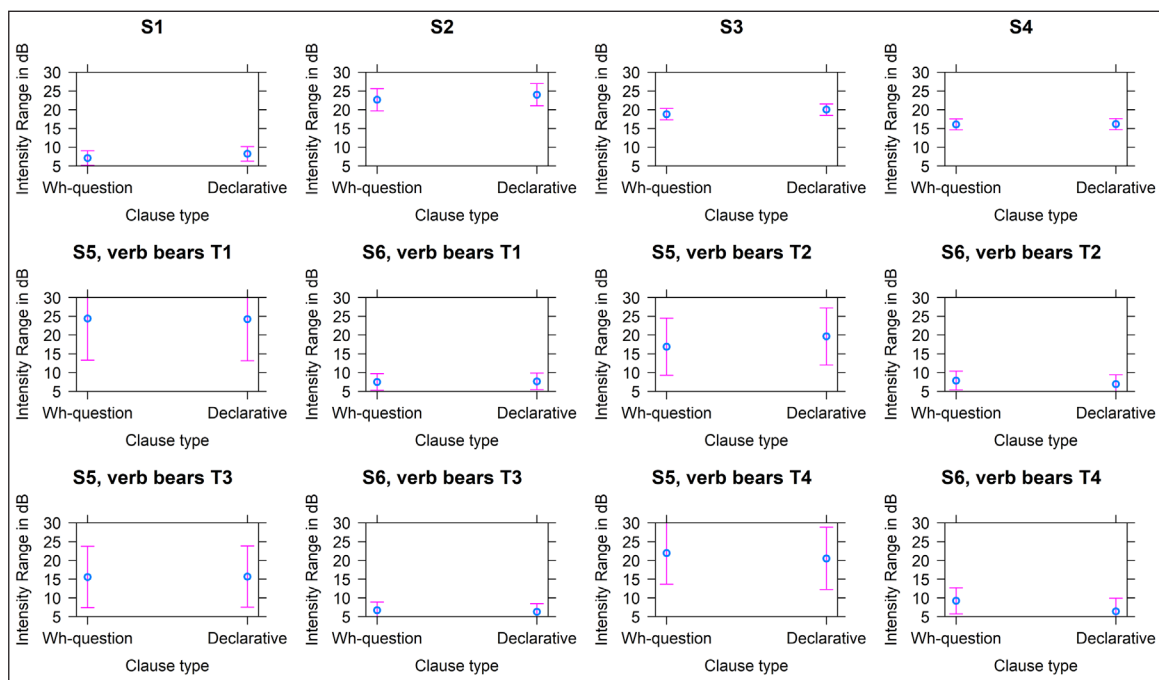


Figure 12: Upper panel: Mean intensity range in dB of syllables S1–S4 in *wh*-questions and declaratives. Middle and lower panel: Mean intensity range in dB for syllables S5 and S6 in *wh*-questions and declaratives.

From the above analysis, we see that all three acoustic parameters, duration, F0, and intensity of the stimuli used for the gating study are in the same direction as the acoustic parameters of the stimuli obtained from the production experiment, although they cannot be exactly the same, as the stimuli of the gating study come from one particular speaker.

Response stimuli. We also prepared two types of response stimuli that appeared written on screen after each audio fragment was heard. The response stimuli consisted of either a *wh*-question or a declarative continuation; the two continuations differed only at the *wh* and its non-*wh*-counterpart. An example of the two sentence continuations for gate-a is given in (14)–(15). For gate-b, the two sentence continuations were identical to the ones for gate-a, but without the adverb (e.g., *zuótiān* ‘yesterday’). For gate-c, the two sentence continuations were identical to the ones for gate-b, but without the verb and the perfective marker *le* (e.g., *bāo-le* ‘peeled’).

- (14) *wh*-question continuation
zuótiān bāo-le shénme gěi Luó Yīng?
 yesterday peel.PERF what for Luo Ying
 ‘peeled what for Luo Ying yesterday?’
- (15) declarative continuation
zuótiān bāo-le júzi gěi Luó Yīng
 yesterday peel.PERF oranges for Luo Ying
 ‘peeled oranges for Luo Ying yesterday.’

3.1.3. Procedure

The experiment was run using MFC, Praat (Boersma & Weenink, 2017) and proceeded as follows. Participants were seated in front of a computer and were asked to read the instructions that appeared on the computer screen and to press the OK button once they were ready to start with the experiment. The first audio fragment was played via the computer’s loudspeakers 1.0 seconds after clicking on the OK button. While the audio fragment was played, the screen was empty and 0.3 seconds after the fragment’s offset, two possible sentence continuations appeared on the screen. The participants’ task was to select one of the two responses on the basis of the audio fragment they had heard and to confirm their selection by clicking on the OK button. The next audio stimulus was played 1.0 seconds after clicking on the OK button. Participants listened first to the audio fragments of gate-a, then the audio fragments of gate-b, and finally the audio fragments of gate-c. The order of presenting the two sentence continuations was randomized to avoid any presentation bias. We also randomized the order of presenting the audio fragments in every gate to avoid any presentation bias. The experiment took place in a silent room at Tsinghua University in Beijing and lasted approximately 20 minutes.

3.2. Statistical analysis

We used a series of mixed-effect models to analyze the likelihood of the type of continuation (*wh*-questions versus declaratives). All the analyses were run in R using the *lme4* Package (Bates et al., 2015). We first ran null models that included participants’ responses as dependent variable, and participants and items as random factors. We then ran additional models adding as a predictor the clause type intended by the speaker to see whether the model improved. For gate-c we also ran a third model adding a second predictor, namely, the verb tone, to see whether the model improved, and then a fourth model in which the interaction between the two predictors is also included. For gate-a and gate-b, the null models improved when the clause type intended by the speaker was added as a predictor to the model [for gate-a: $\chi^2(1) = 90.68, p < 0.001$, for gate-b: $\chi^2(1) = 212.77, p < 0.001$].

For gate-c the model with both the clause type intended by the speaker and the verb tone as predictors was better than the model with only the clause type intended by the speaker as predictor [$\chi^2(1) = 4.9096, p < 0.05$]. The model fit did not improve when we added the interaction of clause type \times verb tone [$\chi^2(1) = 1.1808, p > 0.05$]. Lastly, we examined the effect of gates on participants' responses running a model that included participants' responses as dependent variable, clause type and gates as predictors, and participants and items as random factors.

3.3. Results

We obtained a total of 4320 responses (3 gates \times 40 stimuli \times 36 participants). In general, participants were successful in identifying the type of the clause that was intended by the speaker; see **Figures 13–14**. Moreover, there was a significant association between the type of the clause that was intended by the speaker and the participants' responses [gate-a: $\chi^2(1) = 84.31, p < 0.001$, gate-b: $\chi^2(1) = 189.09, p < 0.001$, gate-c: T1 = $\chi^2(1) = 57.66, p < 0.001$, T2 = $\chi^2(1) = 39.56, p < 0.001$, T3 = $\chi^2(1) = 18.86, p < 0.001$, T4 = $\chi^2(1) = 80.31, p < 0.001$]. As shown in the left panel of **Figure 13**, in gate-a, when the speaker intended a *wh*-question, the listeners chose a *wh*-question more often than a declarative continuation. Similarly, the listeners chose a declarative continuation more often than a *wh*-question continuation when the speaker intended a declarative.

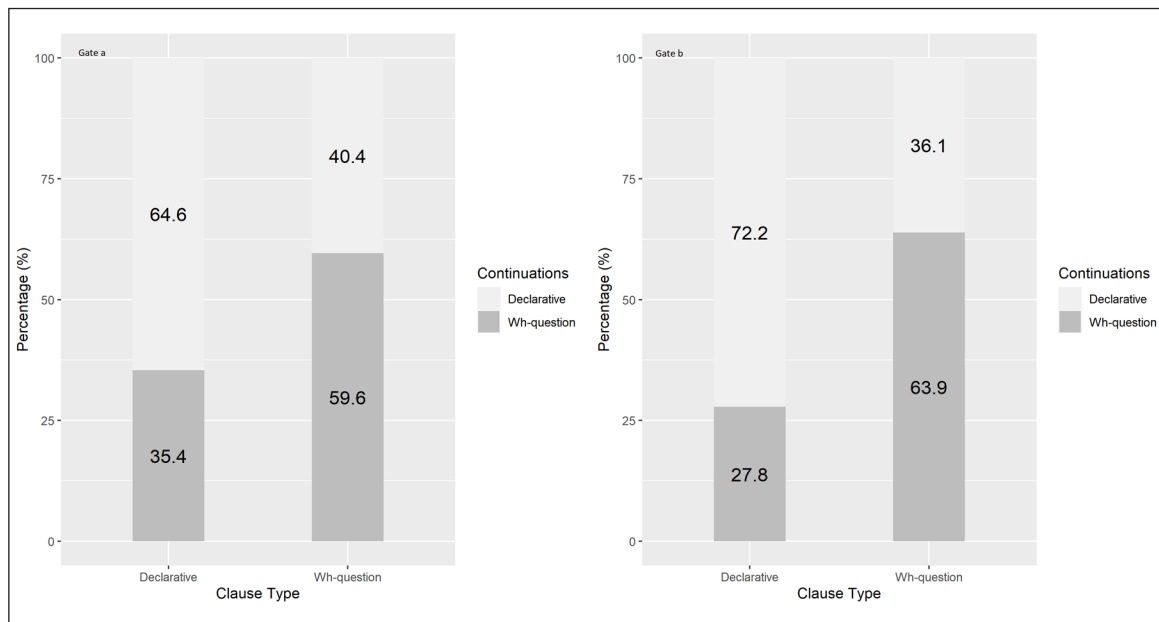


Figure 13: Listeners' responses in percentage (%) to audio stimuli of gate-a (e.g., *Máo Kē* 'Mao Ke') and gate-b (e.g., *Máo Kē zuótiān* 'Mao Ke yesterday') in the left and right panel respectively.

The right panel of **Figure 13** presents participants' responses to gate-b. As shown in this figure, listeners chose a *wh*-question continuation more often than a declarative continuation for audio fragments that originated from *wh*-questions. Likewise, for audio fragments that originated from declaratives, listeners chose a declarative continuation more often than a *wh*-question.

Figure 14 presents participants' responses to gate-c broken down by the tone of the verb. In general, participants were able to identify the type of the clause that was intended by the speaker when the verb bore T1, T2, and T4. As shown in the left upper panel in **Figure 14**,

when the verb bore T1, participants chose a *wh*-question continuation for audio fragments that originated from *wh*-questions more often than a declarative continuation. Likewise, they chose a declarative continuation more often than a *wh*-question continuation for audio fragments that originated from declaratives. After listening to audio fragments of gate-c with T2 on the verb, participants responded in a similar way; audio fragments that originated from *wh*-questions triggered more *wh*-continuation responses, while fragments that originated from declaratives received more declarative continuation responses. Audio fragments with T4 on the verb elicited similar responses by the participants. The distribution of the participants' responses was different when the verb bore T3. After listening to audio fragments of gate-c with T3 on the verb, listeners had difficulty in identifying the type of the clause that was intended by the speaker. As shown in the lower left panel in **Figure 14**, participants chose a declarative continuation 66.7% of the times when the intended clause type was a *wh*-question.

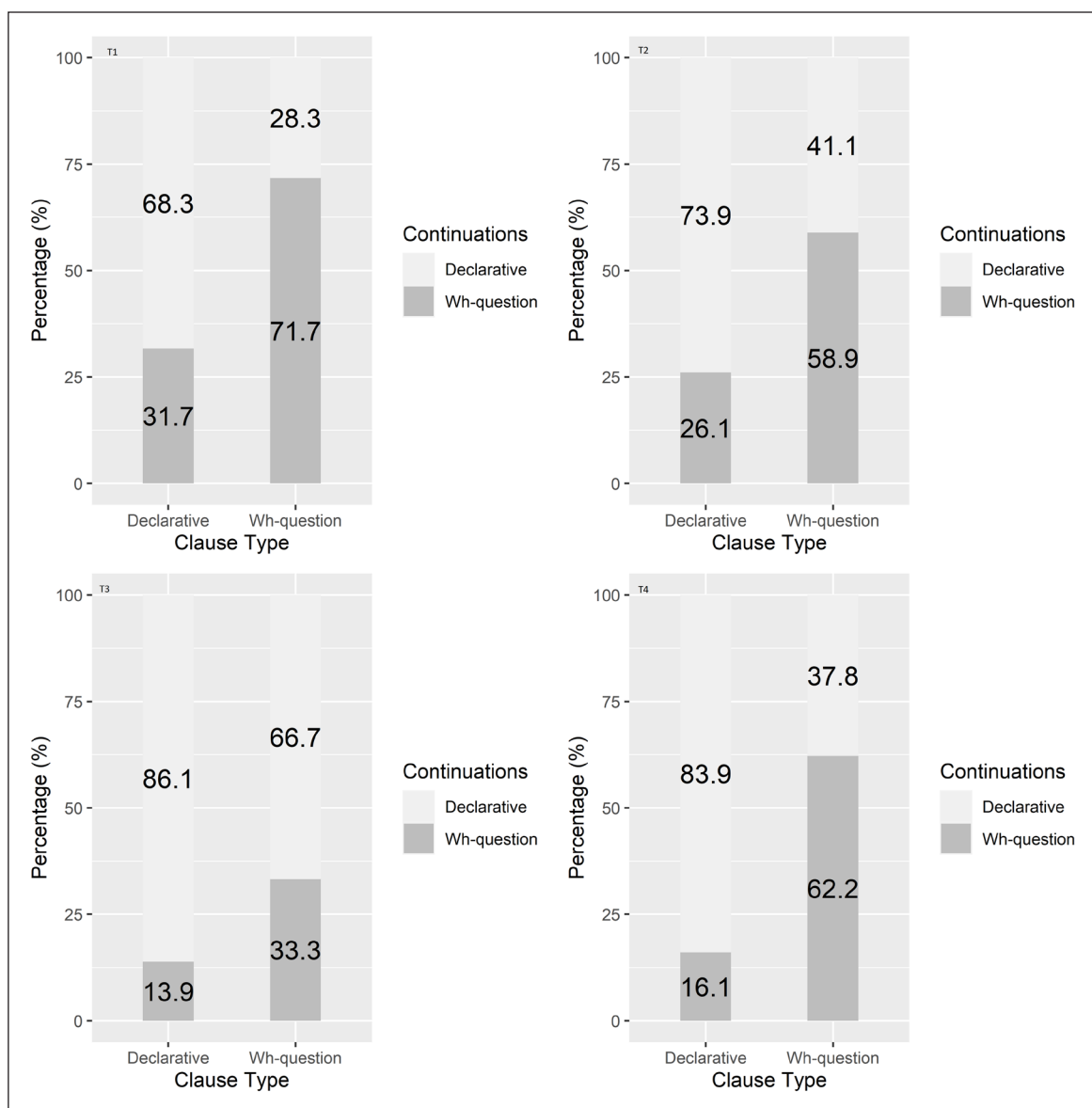


Figure 14: Listeners' responses in percentage (%) to stimuli of gate-c broken by tone on the verb. The upper left panel presents listeners' responses when the verb bore T1, the upper right panel for verb = T2, while the lower left panel for verb = T3 and the lower right panel for verb = T4.

Moreover, linear mixed-effect models showed that the type of the clause that was intended by the speaker affected participants' responses in gate-a, gate-b, and gate-c [Gate-a: Estimates = 1.0579, $SE = 0.1135$, $p < 0.001$; gate-b: Estimates = 1.7033, $SE = 0.1239$, $p < 0.001$; gate-c: Estimates = 1.2020, $SE = 0.2741$, $p < 0.001$]. Furthermore, we examined whether the tone of the verb affected participants' responses. Our results showed that T3 on the verb affected significantly participants' responses. In particular, after hearing an audio fragment with T3 on the verb, participants were less likely to choose a *wh*-question continuation than after hearing an audio fragment with one of the other tones on the verb (T3 versus T1: [Estimates = -1.7548 , $SE = 0.3362$, $p < 0.001$]; T3 versus T2: [Estimates = -1.1392 , $SE = 0.3283$, $p < 0.001$]; T3 versus T4: [Estimates = -1.2967 , $SE = 0.3303$, $p < 0.001$]; see **Figure 15**. Moreover, this figure shows that interaction between the clause type that was intended by the speaker and T3 differed significantly from the interaction between the clause type that was intended by the speaker and T4. Furthermore, as shown in **Figure 15** (see left panel) when the intended clause type is a *wh*-question, the participants were more likely to choose a *wh*-continuation after listening to an audio fragment with T4, than after listening to an audio fragment with T3.

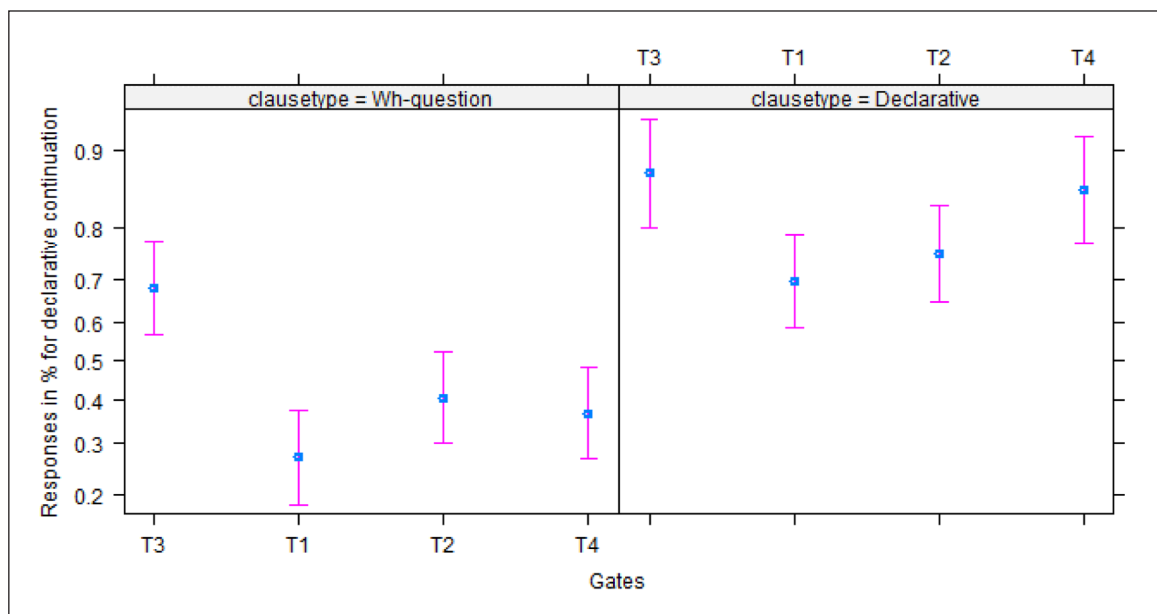


Figure 15: Effects of clause type and verb tone on participants' responses.

Furthermore, the results of a linear mixed-effects model analysis showed that the distribution of participants' responses did not differ between gate-a and gate-b [Estimates = -0.194 , $SE = 0.111$, $p > 0.05$]; see **Figure 16**. Between gate-b and gate-c the results differed depending on the tone on the verb. When the verb bore T1, the participants were more likely to select the question continuation [Estimates = -0.595 , $SE = 0.197$, $p < 0.05$], while they were less likely to do so when the verb bore T3 [Estimates = 1.160, $SE = 0.193$, $p < 0.001$]. The distribution of participants' responses did not differ significantly between gate-b and gate-c when the verb bore T2 [Estimates = 0.210, $SE = 0.187$, $p > 0.05$] or T4 [Estimates = 0.051, $SE = 0.188$, $p > 0.05$].

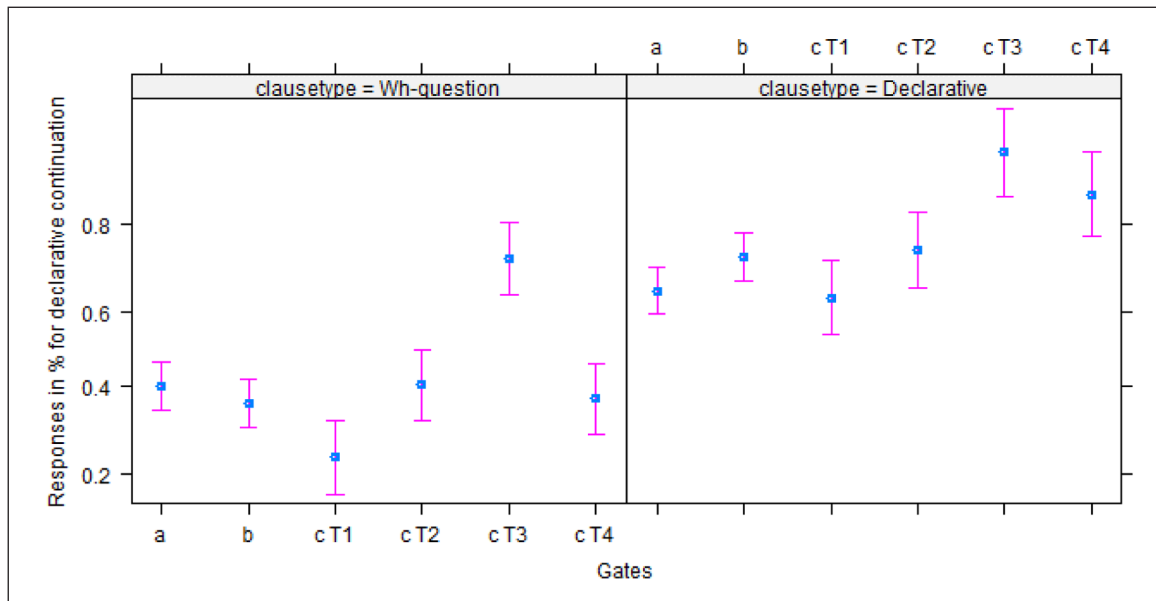


Figure 16: Effects of clause type and gate on participants' responses.

4. Discussion

The current study is the first to offer an in-depth comparison of the pre-*wh* region of *wh*-in situ questions and their declarative counterparts. The set-up of the study allowed us to examine the prosodic differences between these two types of sentences in much more detail than in previous studies. It also allowed us to compare the interaction between tones and intonation in this region. We will first summarize the overall results of the study in relation to the previous literature. Then, in Subsection 4.2, we will discuss the interaction of tone and prosody in clause type anticipation in the perception experiment.

4.1. Overall results

The results of our production study confirm the observation of Lee (2005) concerning overall pitch, though based on a much larger sample of utterances. In-situ *wh*-questions present a higher overall pitch than corresponding declarative clauses that function as statements. Moreover, our data show that the pre-*wh* region of a *wh*-in situ question is characterized by a reduced pitch range when the verb bears T1 or T4 as compared to that same region in a corresponding declarative sentence, while an expanded pitch range can be observed on the *wh*-constituent. As we will show below, our data also show that tones interact with overall pitch and F0-range. Besides pitch, we also found effects of duration and intensity; these effects have not been previously reported in the literature. The mean syllable duration in *wh*-in-situ questions is significantly shorter than in corresponding declaratives. As for intensity, the syllable corresponding to the last syllable of the adverb (S4) presented a significantly higher intensity range in the in-situ *wh*-questions than in the corresponding declaratives.

Our perception study showed that prosodic cues in the speech signal permit listeners to discriminate between *wh*-questions and declarative statements above chance level. In what follows, we will discuss the relation between the cues present in the audio fragments that the participants heard and the detailed results of the gating task in order to get a better idea of the effects of different cues, such as pitch, duration, and intensity. We will first limit ourselves to gate-a (S1–S2) and gate-b (S1–S4), that is, the gates in which there was no

variation with respect to tones among the items, and to the global results of gate-c (S1–S6). In the next section, we will turn to the effects of the different tones on the verb in S5.

The audio fragments that were presented to the listeners at gate-a offered two types of cues for distinguishing between *wh*-questions and declaratives. In the first place, the F0 maximum (H) on S2 (T1) was significantly higher for questions than for declaratives. In the second place, the duration of both syllables was significantly shorter in questions than in declaratives. There were no differences in intensity. At gate-b, the F0 maximum (H) on S4 was significantly higher for questions than for declaratives, while S3 and S4 were significantly shorter in questions. To some extent, the listeners did not have much more information than at gate-a. These descriptive differences are in line with the fact that the results for gate-a and gate-b were similar, suggesting that the cues used by the listeners were most likely already contained in the first two syllables. The extra duration information on S3 and S4 did not lead to better discrimination, suggesting that duration information in S3/S4 was either not used, or the information in the first two syllables was enough to perceive a higher speech rate. At gate-c, the results differed per verb tone. In the next section, we will consider the interaction between intonation and tones in more detail.

4.2. Interaction between tone and intonation

The interaction between tone and intonation in Mandarin has been mainly discussed in declarative questions in comparison with their string identical declaratives. As mentioned in Section 2, the most striking F0 difference between declarative questions and ordinary declaratives lies in the final syllable. That is, the final syllable in declarative questions is normally higher in F0, showing a rising contour, as compared with that in declaratives. Since F0 is not only (one of) the most important acoustic correlates of intonation, but also the most important acoustic correlate of tones in Mandarin, many scholars start to investigate the interaction between intonation and the rising tone (T2) and the falling tone (T4). In particular, the question addressed is whether these tones at the end of the sentence affect the identification of intonations/clause types, or vice versa, whether the intonation affects the identification of tones (Yuan, 2006; Liu, Chen, & Schiller, 2016; Ma, Giocca, & Whitehill, 2010). However, no literature has addressed the interaction of tone 3 and intonation.

The consistent conclusion of these studies is that there is an interaction between tone and intonation especially when the sentence ends in a T2. For instance, Yuan (2011) found that in Mandarin, declarative questions ending with T4 (falling tone) were easier to identify than declarative questions ending with T2 (rising tone). Liu et al. (2016) found that Mandarin listeners can distinguish between declarative-question intonation and declarative intonation when the intonation is associated with a final T4, but fail to do so when the intonation is associated with a final T2.

Even though our study was not designed to study the interaction between tones and intonation, and certainly not in comparison with results from previous studies, our results from both the production experiment and the audio perception study show that the tone on the verbs interacted with other factors. In particular, the items containing a verb with T3 behaved differently from the ones with T1, T2, and T4, even though we also observed effects of T1 and T4.

In the production experiment, the most important effects are found for pitch. Whereas T1, T2, and T4 gave rise to a reduced F0-range and a higher overall pitch in the pre-*wh* region of the *wh*-in situ question as compared to the corresponding region in the declarative (see the general pattern described in the previous section), these effects were not observed for T3. When looking at specific measurement points, the final particle *le* (S6) showed

different patterns for T1 and T4. Whereas we found overall higher pitch for T2 and T3 in questions than in declaratives, T1 and T4 give rise to a lower F0 maximum H and a higher F0 minimum L for this syllable. As the data in **Figure 4** show, this lower pitch maximum in the *wh*-in situ questions results in a clearer difference between the two clause types: In the declaratives there is a F0 rise between the verb in S5 and the fall on the verbal particle *le* in S6, while there is a continuous fall in the *wh*-questions. Besides pitch, intensity also interplayed with the verb tone. T4 behaved differently from the other tones, as it exhibited a significantly higher intensity range in *wh*-questions, while *wh*-questions with a verb that bore T3 were more similar to the corresponding region in declaratives in terms of their overall pitch, their pitch range, and intensity.

Turning to the perception experiment, tone played a role at gate-c as this gate also contained the verb (bearing tones T1–T4) and the perfective particle *le*. After hearing audio fragments of gate-c, a difference was observed for T3 as opposed to T1, T2, and T4. In these latter cases, the listeners identified the intended clause type of both questions and declaratives above chance level. However, after hearing fragments with a T3 verb, the listeners correctly identified declarative sentences (86.11% correct responses), but they confused *wh*-questions with declaratives (33% correct responses). This suggests that the prominence of the low-dipping tone, Tone 3, affects the perception of the other prosodic cues, and seems to overrule prosodic information provided in S1–S4. Moreover, the clear pitch difference on the particle *le* did not lead to correct identification of the clause type.

An alternative interpretation of the results of the perception experiment is that the female speaker that we chose to use in the perception experiment realizes T3 in such a way that confuses listeners with respect to the identification of the clause type; this confusion appears at the verb position. Had we used another speaker whose production data were more often significantly different between the two clause types, then the results could be different. Such hypothetical results would imply that listeners are particularly sensitive to the prosodic cues they hear in the corresponding gates. Alternatively, we could manipulate duration, F0, and intensity and examine their respective contribution to clause type identification. Another related issue is gender; in this perception experiment we chose to use a female speaker. Had we used a male speaker, then the results could have been different. At this point we do not know whether gender is a relevant factor for clause type anticipation.

As for the items with T1, our data show that these were more often interpreted as questions, suggesting that the high level tone (T1) facilitated the correct identification of questions (71.7% of correct responses). T1 and T4 were the only tones in which the maximum pitch on the verb was higher in the fragments used for the audio perception study, but this only helped question identification with T1. For T4 and for T2 no significant difference was found with respect to gate-b.

5. Conclusion

Both of the research questions that we posed in Section 1 have positive answers: Speakers of Mandarin *do* prosodically mark the clause type in the pre-*wh*-word contour; and listeners do use this prosodic marking to determine the clause type before reaching the *wh*-word or its non-*wh*-counterpart. As the *wh*-in-situ questions investigated in our study are not biased questions, we suggest that the prosodic markings that we identified reflect the clause type of these questions. We have seen that though listeners do use the prosodic marking in determining the clause-type, the tonal properties of the stimuli must also be taken into consideration.

Additional Files

The additional files for this article can be found as follows:

- **Appendix A.** List of stimuli of the production experiment. DOI: <https://doi.org/10.5334/labphon.169.s1>
- **Appendix B.** A detailed overview of the results of a series of linear mixed effects models that examined the effect of the type of the clause on duration, speech rate, F0, and intensity. DOI: <https://doi.org/10.5334/labphon.169.s2>

Acknowledgements

The research reported here was funded by the Dutch Research Council (NWO) via the project Understanding Questions (360-70-480). We would like to thank three anonymous reviewers and the journal's editors for their insightful comments on the paper. We would also like to thank our team members Aliza Glasbergen-Plas and Leticia Pablos for comments and discussion. We thank Roger Luo, our assistant for segmenting the production data. We would also like to thank Xiaolu Yang at Tsinghua University, Beijing for assistance in conducting the experiments. At last we would like to thank all our participants.

Competing Interests

The authors have no competing interests to declare.

References

- Agresti, A.** (2002). *Categorical data analysis*. 2nd Edition, Hoboken, NJ: Wiley. DOI: <https://doi.org/10.1002/0471249688>
- Altmann, G., & Kamide, Y.** (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*, 247–264. DOI: [https://doi.org/10.1016/S0010-0277\(99\)00059-1](https://doi.org/10.1016/S0010-0277(99)00059-1)
- Altmann, G., & Kamide, Y.** (2007). The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language*, *57*, 502–518. DOI: <https://doi.org/10.1016/j.jml.2006.12.004>
- Arai, M., & Keller, F.** (2013). The use of verb-specific information for prediction in sentence processing. *Language and Cognitive Processes*, *28*(4), 525–560. DOI: <https://doi.org/10.1080/01690965.2012.658072>
- Baltazani, M., Kainada, E., Lengeris, A., & Nikolaidis, K.** (2015). The prenuclear field matters: Questions and statements in Standard Modern Greek. *Proceedings of the 18th International Congress of Phonetic Sciences*, Glasgow, Scotland UK.
- Bates, D., Mächler, M., Bolker, B., & Walker, S.** (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48. DOI: <https://doi.org/10.18637/jss.v067.i01>
- Boersma, P., & Weenink, D.** (2017). *Praat: Doing phonetics by computer* [computer program] version 6.0.32. <http://www.praat.org/>.
- Brown, M., Salverda, A. P., Dilley, L. C., & Tanenhaus, M. K.** (2011). Expectations from preceding prosody influence segmentation in online sentence processing. *Psychonomic Bulletin & Review*, *18*(6), 1189–1196. DOI: <https://doi.org/10.3758/s13423-011-0167-9>
- Chen, S. H.** (2005). The effects of tones on speaking frequency and intensity ranges in Mandarin and Min dialects. *The Journal of the Acoustical Society of America*, *117*, 3225. DOI: <https://doi.org/10.1121/1.1872312>
- Chen, Y.** (2010). Post-focus F0 compression – Now you see it, now you don't. *Journal of Phonetics*, *38*, 517–525. DOI: <https://doi.org/10.1016/j.wocn.2010.06.004>

- Chen, Y., & Gussenhoven, C.** (2008). Emphasis and tonal implementation in Standard Chinese. *Journal of Phonetics*, *36*, 724–746. DOI: <https://doi.org/10.1016/j.wocn.2008.06.003>
- Cheng, C.C.** (1968). English stress and Chinese tones in Chinese sentences. *Phonetica*, *18*, 77–88. DOI: <https://doi.org/10.1159/000258601>
- Cheng, L. L.-S.** (1991). *On the typology of wh-questions*. Doctoral dissertation. MIT.
- Clark, A.** (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(03), 181–204. DOI: <https://doi.org/10.1017/S0140525X12000477>
- Cooper, S.** (2015). Intonational signalling of sentence type in northern Welsh. *Proceedings of the 18th International Congress of Phonetic Sciences*, Glasgow, Scotland UK.
- De Francis, J.** (1963). *Beginning Chinese*. New Haven: Yale University Press.
- Duanmu, S.** (2000). *The phonology of Standard Chinese*. Oxford: Oxford University Press.
- Duanmu, S.** (2011). Chinese syllable structure. In M. van Oostendorp, C. J. Ewen, E. V. Hume & K. Rice (Eds.), *The Blackwell companion to phonology, Vol. V: Phonology across languages* (pp. 2754–2777). Chichester: Wiley-Blackwell. DOI: <https://doi.org/10.1002/9781444335262.wbctp0115>
- Face, T.** (2007). The role of intonational cues in the perception of declaratives and absolute interrogatives in Castilian Spanish. *Estudios de Fontica Experimental*, *16*, 185–225.
- Friston, K.** (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, *11*(2), 127–138. DOI: <https://doi.org/10.1038/nrn2787>
- Grosjean, F.** (1980). Spoken word recognition and the gating paradigm. *Perception & Psychophysics*, *28*(4), 267–283. DOI: <https://doi.org/10.3758/BF03204386>
- Gunlogson, C.** (2002). Declarative questions. In B. Jackson (Ed.), *SALT XII* (pp. 124–143). Ithaca, NY: Cornell University. DOI: <https://doi.org/10.3765/salt.v12i0.2860>
- Haan, J., van Heuven, V., Pacilly, J., & van Bezooijen, R.** (1997). On the Anatomy of Dutch Question Intonation. In H. de Hoop & J. Coerts (Eds.), *Linguistics in the Netherlands* (pp. 99–110). John Benjamins.
- Heeren, W., Bibyk, S., Gunlogson, C., & Tanenhaus, M.** (2015). Asking or telling – real-time processing of prosodically distinguished questions and statements. *Language and Speech*, *58*(4), 474–501. DOI: <https://doi.org/10.1177/0023830914564452>
- Ho, A. T.** (1977). Intonation variations in a mandarin sentence for three expressions: Interrogative, exclamatory, and declarative. *Phonetica*, *34*, 446–456. DOI: <https://doi.org/10.1159/000259916>
- Huang, C. T. J.** (1982). Move *wh* in a language without *wh*-movement. *Linguistic Review*, *1*(4), 369–416. DOI: <https://doi.org/10.1515/tlir.1982.1.4.369>
- Ito, K., & Speer, S. R.** (2008). Anticipatory effect of intonation: Eye movements during instructed visual search. *Journal of Memory and Language*, *58*, 541–73. DOI: <https://doi.org/10.1016/j.jml.2007.06.013>
- Kamide, Y., Scheepers, C., & Altmann, G. T. M.** (2003). Integration of syntactic and semantic information in predictive processing: Cross-linguistic evidence from German and English. *Journal of Psycholinguistic Research*, *32*, 37–55. DOI: <https://doi.org/10.1023/A:1021933015362>
- Lahiri, A., & Marslen-Wilson, W.** (1991). The mental representation of lexical form: A phonological approach to the recognition lexicon. *Cognition*, *38*, 254–294. DOI: [https://doi.org/10.1016/0010-0277\(91\)90008-R](https://doi.org/10.1016/0010-0277(91)90008-R)
- Lee, O.** (2005). *The prosody of questions in Beijing Mandarin*. Doctoral dissertation. Ohio State University.
- Li, A.** (2002). Chinese Prosody and Prosodic Labeling of Spontaneous Speech. *Proceedings of Speech Prosody 2002*, Aix en Provence, France, 39–46.

- Li, B. (2006). *Chinese final particles and the syntax of the periphery*. Doctoral dissertation. Leiden University.
- Li, C., & Thompson, S. (1981). *Mandarin Chinese: A Function Reference Grammar*. Los Angeles, California: University of California Press.
- Liu, F. (2009). *Intonation systems of Mandarin and English: A functional approach*. Doctoral Dissertation. University of Chicago.
- Liu, M., Chen, Y., & Schiller, N. O. (2016). Online processing of tone and intonation in Mandarin: Evidence from ERPs. *Neuropsychologia*, 91, 307–317. DOI: <https://doi.org/10.1016/j.neuropsychologia.2016.08.025>
- Ouyang, I. C., & Kaiser, E. (2013). Prosody and information structure in a tone language: An investigation of Mandarin Chinese. *Language, Cognition and Neuroscience*, 30, Nos 1–2, 57–72. DOI: <https://doi.org/10.1080/01690965.2013.805795>
- Petrone, C., & D’Imperio, M. (2008). Tonal structure and constituency in Neapolitan Italian: Evidence for the accentual phrase in statements and questions. *Proceedings of Speech Prosody 2008, Campinas, Brazil*, 301–304.
- Petrone, C., & Niebuhr, O. (2014). On the Intonation of German Intonation Questions: The Role of the Prenuclear Region. *Language and Speech*, 57, 108–146. DOI: <https://doi.org/10.1177/0023830913495651>
- R Core Team. (2017). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from <https://www.R-project.org/>
- Shen, X. (1990). *The prosody of Mandarin Chinese*. Berkeley: University of California Press.
- Staub, A., & Clifton, J. (2006). Syntactic prediction in language comprehension: Evidence from *either...or*. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32(2), 425–436. DOI: <https://doi.org/10.1037/0278-7393.32.2.425>
- Titze, I. R. (1988). Regulation of vocal power and efficiency by subglottal pressure and glottal width. In O. Fujimura (Ed.), *Vocal Fold Physiology: Voice Production Mechanisms and Functions* (pp. 227–238). New York: Raven.
- Tsao, W.-Y. (1967). Question in Chinese. *Journal of the Chinese Language Teachers’ Association*, 2, 15–26.
- van de Weijer, J., & Sloos, M. (2014). The four tones of Mandarin Chinese. *Linguistics in the Netherlands 2014*, 180–191. DOI: <https://doi.org/10.1075/avt.31.13wei>
- van Heuven, V., & Haan, J. (2002). Temporal distribution of interrogativity markers in Dutch: A perceptual study. In C. Gussenhoven & N. Warner (Eds.), *Papers in Laboratory Phonology 7* (pp. 61–86). Mouton de Gruyter. DOI: <https://doi.org/10.1515/9783110197105.61>
- Xu, Y. (1999). Effects of tone and focus on the formation and alignment of F0 contours. *Journal of Phonetics*, 27, 55–105. DOI: <https://doi.org/10.1006/jpho.1999.0086>
- Yuan, J. (2006). Mechanisms of question intonation in Mandarin. In Q. Huo, et al. (Eds.). *ISCSLP 2006, LNAI, 4274*, 19–30. DOI: https://doi.org/10.1007/11939993_7
- Yuan, J. (2011). Perception of intonation in Mandarin Chinese. *Journal of Acoustic Society of America*, 130(6), 4063–4069. DOI: <https://doi.org/10.1121/1.3651818>

How to cite this article: Gryllia, S., Doetjes, J. S., Yang, Y., & Cheng, L. L. 2020 Prosody, clause typing, and *wh*-in-situ: Evidence from Mandarin. *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 11(1):19, pp.1–30. DOI: <https://doi.org/10.5334/labphon.169>

Submitted: 04 October 2018

Accepted: 23 June 2020

Published: 02 November 2020

Copyright: © 2020 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.



Laboratory Phonology: Journal of the Association for Laboratory Phonology is a peer-reviewed open access journal published by Ubiquity Press.

OPEN ACCESS