The two sides of Wh-indeterminates in Mandarin : a prosodic and processing account

Yang, Y.

Chapter 3 Clause Type Anticipation Based on Prosody — An Audio-perception and Gating study

3.1 Introduction

In the previous chapter we investigate how prosody differentiates clause types from the perspective of the speakers. In the current chapter, we explore a related topic from the perspective of the listeners, namely, clause type anticipation. Before scrutinizing clause type anticipation, we first briefly introduce the anticipation in spoken sentences. Our daily communication involves anticipation in the oral speech processing and here anticipation often refers to predicting how the discourse of a speaker will evolve (Seeber, 2001). To be precise, spoken sentences proceed in an incremental manner and when a sentence unfolds over time, listeners interpret the information and formulate a prognosis of how the sentence may further unfold. Generally speaking, listeners can process incoming information rapidly to dissolve ambiguities and predict the most plausible interpretation from the sentence (Braun & Chen, 2012). For example, listeners can predict upcoming referents by utilizing intonation cues such as accent (see Braun & Chen, 2012 for details). To make predictions/anticipations, listeners use all kinds of linguistic knowledge and rely on every cue they may get, in combination of information from syntax, morphology, discourse-semantics and prosody/intonation (Kohn & Kalina, 1996; Braun & Chen, 2012).

With respect to anticipation and prosody, a few studies so far attempted to establish a link between the two. These studies mainly focus on the role of pitch accent/prominence in reference resolution or in information structure interpretation (Terken & Hirschberg, 1994; Dahan, Tanenhaus & Chambers, 2002; Weber, Braun & Crocker, 2006; Chen, den Os & de Ruiter, 2007, among others), how the stress pattern of the sentence can help to anticipate upcoming stressed syllables and the end of the speech sequence (Shields, McHugh & Martin, 1974; Buxton, 1983), or how listeners can use prosodic information to predict upcoming syntactic structure or to resolve syntactic ambiguity (Beach, Katz & Skowronski, 1996; Kjelgarrd & Speer, 1999; Carlson, Clifton & Frazier, 2001; Snedeker & Trueswell, 2003, among others). Fewer studies have investigated the question of whether clause type (question or declarative) can be anticipated by utilizing prosodic cues only, that is, whether the sentence unfolding to a question or a declarative can be predicted based on pure prosodic information. Among the limited studies on clause type anticipation, the authors examined the role of pitch accent in predicting the correct clause type (yes-no question and declarative) in Castilian Spanish (Face, 2004), the role of downstepped pitch (in the middle of the sentence) in predicting French yes-no questions (Vion & Colas, 2006), the contribution of pre-nuclear pitch accent properties in predicting yes-no questions and declaratives in Northern standard German (Petrone & Niebuhr, 2014), the intonation contours in anticipating clause types in English (e.g. question 'Want a candy?' and declarative 'Want a candy.') (Heeren, Bibyk, Gunlogson & Tanenhaus, 2015), and the prosodic cues (i.e. high F0

onset) in the pre-*wh*-word region in predicting in-situ *wh*-questions and declaratives in Persian.

For a tonal language like Mandarin Chinese, however, little is known about whether and when clause type anticipation takes place based on prosodic cues only. The consensus so far is that in an identical string of Mandarin yes-no questions[14] and declaratives, the sentence final syllable plays the most important role in identifying clause types as "question intonation has the highest prosodic strengths (high F0 curve) in the sentence final syllables" (Yuan, 2004, 2006). It is worth noting that Mandarin Chinese offers an ideal case of investigating clause type anticipation based on prosody, as we can easily find identical strings, not only yes-no questions and their string identical declarative counterparts, but also *wh*-questions and their declarative counterparts. We will spell out below why Mandarin *wh*-questions can be string identical to their declarative counterparts.

As introduced in previous chapters and repeated here, Mandarin is a *wh*-in-situ language in which *wh*-words remain at their base position just as their declarative counterparts do, as illustrated in (1a-b).

(1) a. 罗薇 昨天 买了 什么？ [*wh*-question]
Lúo Wēi zúotiān mǎi-le shénme?
Luo Wei yesterday buy-PERF what
'What did Luo Wei buy yesterday?'

b. 罗薇 昨天 买了 提子。 [declarative]
Lúo Wēi zúotiān mǎi-le tízi.
Luo Wei yesterday buy-PERF grapes
'Luo Wei yesterday bought grapes.'

Moreover, Mandarin *wh*-words like *shénme* are not only in-situ but also known as *wh*-indeterminates, which can have various interpretations (including a number of non-interrogative interpretations) depending on the context and licensors as in Japanese and Korean (Huang, 1982; Cheng, 1991; Li, 1992; Lin, 1998). When *diǎnr* 'a little' appears in front of a *wh*-word, it can have both an interrogative and a declarative interpretation, as illustrated in (2). (2a) is a declarative sentence and the *wh*-word *shénme* is interpreted as an indefinite, meaning 'something'. (2b) is its corresponding *wh*-question. (2a) and (2b) are string identical, but differ in their interpretations.

(2) a. 张三 买了 点儿 什么。 [*wh*-declarative]
Zhāng Sān mǎi-le diǎnr shénme.
Zhang San buy-PERF a.little SHENME
'Zhang San bought a little of something.'

[14] Mandarin yes-no question can have a sentence-final question particle *ma* but *ma* is optionally used, for example *Zhāngsān chī mǐfàn (ma)?* ('Does Zhangsan eat rice?'); when *ma* is not used, the yes-no question is a string identical to its declarative counterpart *Zhāngsān chī mǐfàn.* ('Zhangsan eats rice.').

b. 张三 买了 点儿 什么？ [*wh*-question]
Zhāng Sān mǎi-le diǎnr shénme?
Zhang San buy-PERF a.little SHENME
'What did Zhang San buy (a little of)?'

As reported in the production experiment in Chapter 2, the identical strings of *wh*-questions and *wh*-declaratives have different prosodic markings as early as in the clause initial (i.e. subject position). A question thus arises: can listeners differentiate and anticipate *wh*-questions and *wh*-declaratives by utilizing prosodic cues?

Though no studies have investigated the clause type anticipation of *wh*-declaratives and *wh*-questions so far, we take note of an audio-gating study on Mandarin *wh*-questions and their declarative counterparts with a non-interrogative noun phrase (Gryllia, Yang, Doetjes & Cheng, 2016). As shown in (1a-b), though *wh*-questions and their declarative counterparts are not entirely string identical, they are nonetheless identical up to the *wh*-word / noun phrase. To investigate whether listeners can anticipate clause types in the pre-*wh*-word region, Gryllia et al. (2016) conducted an audio-gating experiment on *wh*-questions and declaratives. They reported that in the beginning of the utterance such as the subject position, listeners can already have a preference towards the correct clause type that was intended by the speaker based on the prosodic cues. The results are insightful in that they indicate that clause type anticipation in Mandarin does take place early. Nevertheless, the question of whether the same anticipation can be found in the string identical case of *wh*-questions and *wh*-declaratives remains to be investigated.

In the current study we address the following questions: 1) Can listeners perceive the prosodic differences between *wh*-questions and *wh*-declaratives as in (2) and make use of them to differentiate the two clause types? 2) If yes, at which point can they perceive them and predict the clause types? In other words, how early does the clause type anticipation take place?

The chapter is organized as follows. In section 3.2, I present the results of a perception experiment that directly address question 1). Section 3.3 presents a series of audio-gating experiments based on *wh*-questions and *wh*-declaratives, the results of which directly address question 2). Section 3.4 is a summary and conclusion.

3.2 Perception experiment

In this section, we tried to tackle our first research question, namely, "Can listeners perceive the prosodic differences between *wh*-questions and *wh*-declaratives and make use of them to differentiate the two clause types?", by designing a perception experiment on the two types of *wh*-sentences. Like in a dialogue, after hearing each *wh*-sentence, participants were asked to continue the discourse by choosing one of two discourse continuations. Example (3) illustrates one type of *wh*-sentence (the one with *wh*-question prosody) which is auditorily presented and example (4) illustrates the two responses/discourse continuations: (4a) a noun phrase and (4b) a *wh*-question, which are visually presented on the screen. Given that (3) is a sentence with *wh*-question prosody, we predict that participants would choose (4a) as a response if they interpret it correctly as a *wh*-question. On the other hand, if the

audio stimulus consists of a sentence with *wh*-declarative prosody, we expect that participants would choose a *wh*-question in order to get more information to continue the discourse, if they correctly perceive the audio as a *wh*-declarative.

(3) 陶薇 昨天 拿了 点儿 什么 给 刘刚？ [*wh*-question]
TáoWēi zúotiān ná-le diǎnr shénme gěi LíuGāng?
TaoWei yesterday bring-PERF a.little what to LiuGang
'What did Tao Wei bring (a little) to Liu Gang yesterday?'

Discourse continuations on the screen

(4) a. 提子。
Tízi.
Grapes
Grapes.

b. 陶薇 拿了 什么？
TáoWēi ná-le shénme?
TaoWei bring-PERF what
'What did TaoWei bring?'

3.2.1 Participants

Thirty-six native speakers of Beijing Mandarin (16 female, 20 male, $\bar{x}$ age = 19 years old) participated in our experiment and were reimbursed for it. They were students at Tsinghua University coming from the northern part of China. None of them have participated in the production experiment reported in Chapter 2. They did not report any hearing or vision disorders (after correction). Prior to recording, informed written consent was obtained from each participant.

3.2.2 Acoustic stimuli

40 stimuli (20 items × 2 clause types) were selected from the recordings of a female native speaker of Beijing Mandarin (age = 20 years old), see the production study in Chapter 2 for details. These stimuli were chosen for their clearness and the speaker's moderate speech rate. Example (5) illustrates a sample stimulus which can be interpreted either as a *wh*-question or as a *wh*-declarative. Each stimulus consisted of 12 syllables and the stimulus length was constant across clause types and items. Below we reported the acoustic properties of the 40 audio stimuli in detail.

(5) a. 陶薇 昨天 拿了 点儿 什么 给 刘刚？ [*wh*-question]
TáoWēi zúotiān ná-le diǎnr shénme gěi LíuGāng?
T2 T1 T2 T1 T2-T0 T3 T2 T0 T3 T2 T1
TaoWei yesterday bring-PERF a.little what to LiuGang
'What did TaoWei bring (a little) to LiuGang yesterday?'

b. 陶薇 昨天 拿了 点儿 什么 给 刘刚. [*wh*-declarative]
TáoWēi zúotiān ná-le diǎnr shénme gěi LíuGāng.
T2 T1 T2 T1 T2-T0 T3 T2 T0 T3 T2 T1
TaoWei yesterday bring-PERF a.little something to LiuGang
'TaoWei brought a little something to LiuGang yesterday.'

Duration. Figure 1 presents the mean sentence duration in milliseconds (ms) of *wh*-questions and *wh*-declaratives. In general, *wh*-questions ($\bar{x}$ = 2049 ms) are shorter than *wh*-declaratives ($\bar{x}$ = 2197 ms). Figure 2 depicts the mean word duration for *wh*-questions and *wh*-declaratives. As shown in the same Figure, already from the subject, there is a consistent duration difference between the two clause types. *Wh*-questions are shorter than *wh*-declaratives at the subject, verb-*le*, and the preposition phrase (e.g., *gei Liu Gang* 'to/for Liu Gang') as well. A series of linear mixed effects models were run in R using the lmerTest package (Kuznetsova, Brockhoff & Christensen, 2013) with sentence/word duration as a dependent variable, clause type as a fixed-effect factor and item as a random factor. The detailed linear mixed effects results are summarized in Table 1.

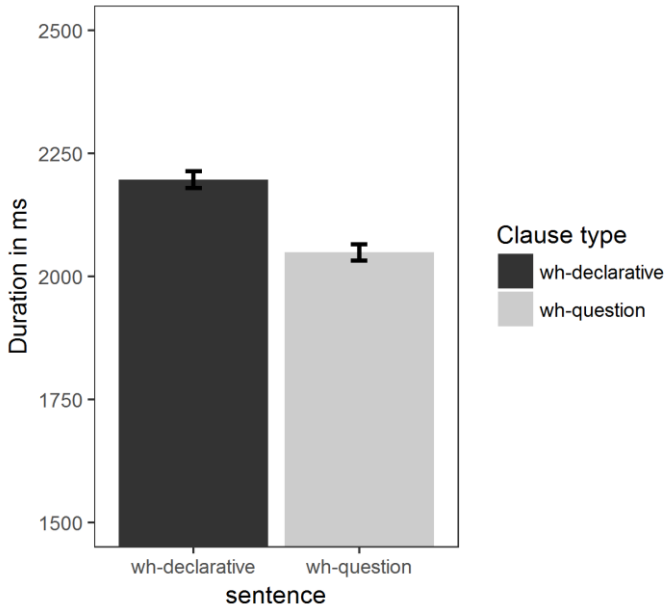

Figure 1. Mean sentence duration across clause types with error bar showing standard error.

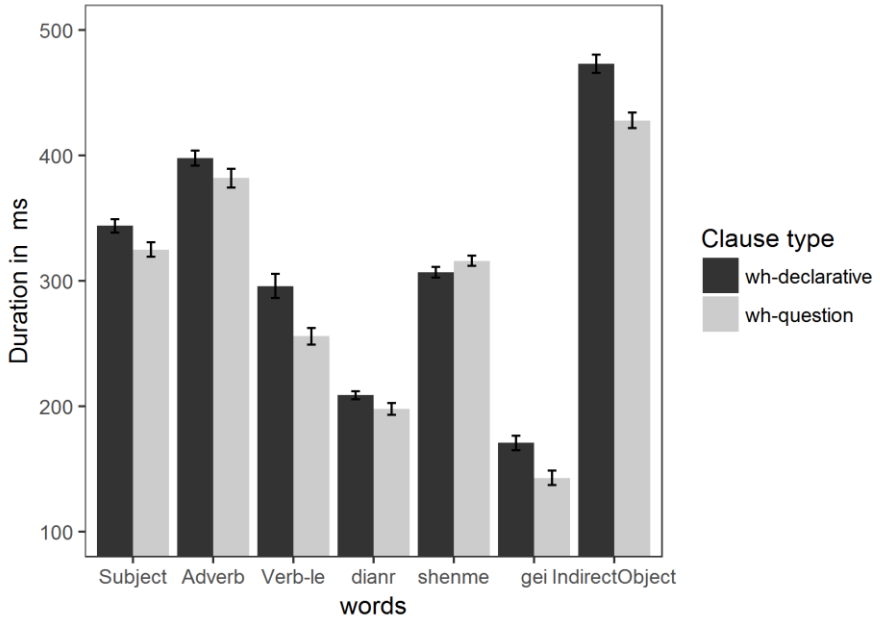

Figure 2. Mean word duration across clause types with error bar showing standard error.

Table 1. Summary of the linear mixed effects models on the duration of each word and the sentence.

	Estimate *β*	Std. Error	*t*-value	*p*-value
subject	18.45	5.36	3.44	< 0.01
verb-*le*	39.45	8.78	4.49	< 0.001
shénme	-9.30	5.97	-1.56	> 0.1
preposition phrase	73.00	9.13	8.00	< 0.001
the whole sentence	148.00	20.00	7.40	< 0.001

F0. As reported in Chapter 2 and repeated here, for the F0 measurement of syllables bearing Tone 1, we measured the F0-maximum (H), while for Tone 3 we measured the F0-minimum (L). For Tone 2 we measured first the F0-minimum and then the F0-maximum (LH), while for Tone 4 we measured first the F0-maximum and then the F0-minimum (HL). For Tone 0 (neutral tone) of the perfective marker *le*, following Li (2002), we measured first the F0-maximum and then the F0-minimum (HL), when the preceding syllable (the verb) bore Tone 1, Tone 2 or Tone 4, while we measured first the F0-minimum and then the F0-maximum (LH), when the preceding syllable bore Tone 3.

Based on the F0 measurement of each syllable (S), the stylized means of F0 curves for *wh*-questions and *wh*-declaratives split per verb tone was given in Figure 3. As we can see, the most striking F0 difference between the two clause types is at

the *wh*-word *shénme*, which shows a steep rise in *wh*-questions but is relatively flat in *wh*-declaratives, and the F0 in *wh*-questions remains higher than that in *wh*-declaratives until the end of sentence. Linear mixed effects model was also run in R with F0 in Hz as a dependent variable and the fixed-effect factor and random factor are the same as in duration models. The detailed linear mixed effects results are summarized in Table 2.

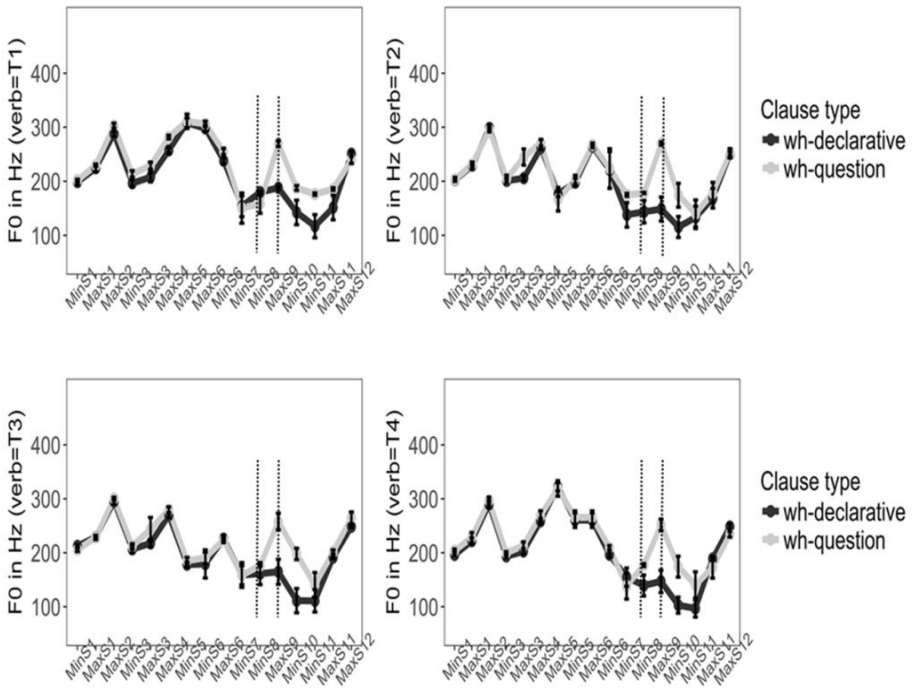

Figure 3. Stylized means of F0 curves across clause types and verb tones with error bar showing standard error; F0 of *shénme* highlighted with dashed lines.

Table 2. Summary of the linear mixed effects models on the F0 measurement between clause types (significant differences were reported).

	Estimate *β*	Std. Error	*t*-value	*p*-value
F0-min adverb (1st syllable)	8.66	2.67	3.24	< 0.01
F0-max adverb (1st syllalbe)	24.65	7.95	3.10	< 0.01
F0-max adverb (2st syllalbe)	14.16	3.88	3.65	< 0.01
F0-min *le*	8.46	3.21	2.64	< 0.05
F0-max *me*	100.60	10.15	9.91	< 0.001
F0-min *gěi*	65.27	11.24	5.81	< 0.001
F0-min indirect object (1st syllable)	34.61	14.23	2.43	< 0.05

3.2.3 Procedure

The perception experiment was conducted in a dim and sound-proof booth in a lab of Tsinghua University in Beijing. Participants were seated in front of a computer where the experiment was running with MFC Praat (Boersma & Weenik, 2016). The procedure was as follows. First, participants read the instructions that appeared on the computer screen and, once they were ready, they pressed OK on the screen to continue. After 1.0 second, the audio stimulus was played (either a *wh*-question or a *wh*-declarative as in (6a-b)). While the audio was played, the screen was blank. 0.3 seconds after the offset of the audio stimulus, two discourse continuations appeared on screen, namely, a noun phrase and a *wh*-question. Participants were instructed to listen to the audio stimulus and then complete the discourse selecting one of the two discourse continuations. After their selection, participants clicked the OK button and 1.0 second after clicking OK, the next audio stimulus was played. The audio stimuli were randomized for each participant to avoid a sequence effect; the two choices on the left or right of the screen were also counterbalanced to avoid any left/right preference among participants. Participants were not forced to make a choice under time pressure. The perception experiment lasted about 10 minutes.

3.2.4 Results

We obtained a total of 1440 responses (40 stimuli × 36 participants). The results showed that participants successfully perceived the clause type that was intended by the speaker and completed the dialogue with the corresponding response/dialogue continuation. On average, when the audio stimulus was a *wh*-question, they correctly perceived it as a *wh*-question and thus chose a noun phrase as response to complete the dialogue at 93.9% of the time. When the audio stimulus was a *wh*-declarative, they correctly perceived it and chose a *wh*-question as a continuation at 95.0% of the time (see Figure 4). The highest accuracy rate per participant was 100%, and the lowest accuracy rate was 70%. The chi-square analysis showed that there was a significant association between clause types intended by the speaker and listeners' responses, $x^2 = 1137.92$ (1), $p < 0.001$.

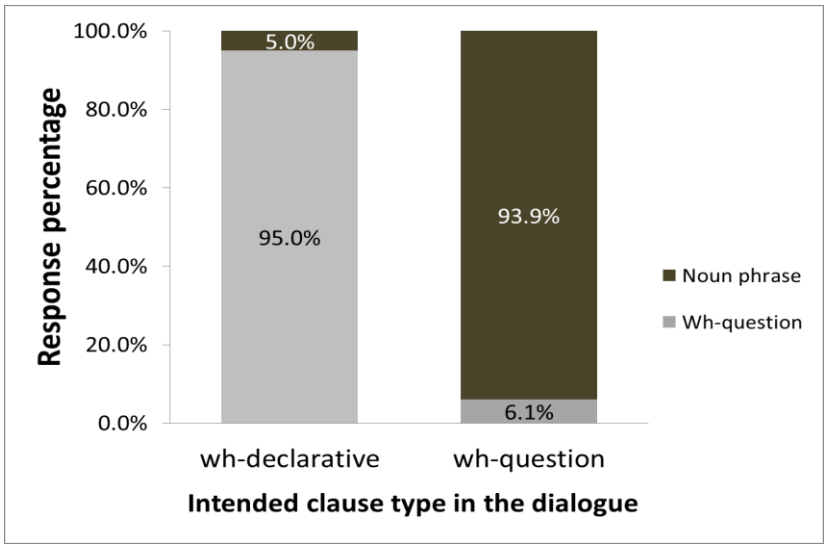

Figure 4. Listeners' responses in percentage (%) in the perception experiment.

3.2.5 Discussion

We conducted the perception experiment to investigate whether the clause types (*wh*-questions and *wh*-declaratives) can be differentiated by listeners. The results demonstrated that participants correctly made use of the prosodic differences between *wh*-questions and *wh*-declaratives to interpret the audio stimuli. When the audio stimulus is a *wh*-question, the participants' accuracy of interpretation is as high as 93.9% and when the audio stimulus is a *wh*-declarative, their accuracy is as high as 95.0%. In short, our experiment provides evidence that the prosodic differences in the string identical case of *wh*-questions and *wh*-declaratives can be perceived and the two clause types can be differentiated by listeners utilizing prosody.

3.3 Audio-gating experiment

As discussed in section 3.1, our second research question is twofold: "At which point can listeners perceive the differences and anticipate the clause types? In other words, how early can listeners perceive the differences?" To answer this research question, we used an audio-gating paradigm where participants listen to audio fragments and choose the corresponding sentence continuations presented on the screen. As we used the same group of participants for the audio-perception experiment and audio-gating experiment, the audio-gating experiment was actually run first to make sure that the information participants heard is incremental and therefore the audio-perception experiment would not affect the results of the audio-gating experiment.

3.3.1 Participants

The same participants from the audio-perception experiment participated in the audio-gating experiment.

3.3.2 Stimuli

The 40 audio stimuli used in the perception experiment (section 3.2.2) were used as a basis for generating audio fragments for the audio-gating experiment. We generated three types of audio fragments that were presented in three gates[15]. Type *a* audio fragment consisted of the subject and was presented in Gate *a*, Type *b* consisted of subject and adverb and was presented in Gate *b*, and Type *d* consisted of subject, adverb, verb plus *le* and *diănr* 'a little', and was presented in Gate *d*. Notice that we included *diănr* in Type *d* together with the verb because *diănr* is described as a clitic to the preceding verb (Shih, 1997; Chen, 2000). For each gate, we also created two kinds of sentence continuations, *wh*-questions and *wh*-declaratives. Examples (6-8) illustrate the fragments used in the audio fragments of three gates and their corresponding sentence continuations visually presented on the screen. As shown, the two kinds of continuations differ only at the sentence final punctuations.

(6) audio Gate *a* — sentence continuations (visually presented)

TáoWēi	-- zúotiān	ná-le	diănr	shénme	gěi	LíuGāng?
TaoWei	yesterday	bring-PERF	a.little	what	to	LiuGang
	-- zúotiān	ná-le	diănr	shénme	gěi	LíuGāng.
	yesterday	bring-PERF	a.little	something	to	LiuGang

(7) audio Gate *b*

TáoWēi zúotiān	-- ná-le	diănr	shénme	gěi	LíuGāng?
TaoWei yesterday	bring-PERF	a.little	what	to	LiuGang
	-- ná-le	diănr	shénme	gěi	LíuGāng.
	bring-PERF	a.little	something	to	LiuGang

(8) audio Gate *d*

TáoWēi zúotiān	ná-le	diănr	-- shénme	gěi	LíuGāng?
TaoWei yesterday	bring- PERF	a.little	what	to	LiuGang
			-- shénme	gěi	LíuGāng.
			Something	to	LiuGang

This resulted in a total of 40 audio fragments of each gate and 120 in total. The detailed acoustic properties can be found in section 3.2.2.

[15] The audio-gating experiment also includes another gate, gate *c*, which is designed for different research questions and is reported in Chapter 4.

3.3.3 Procedure

Similar to the perception experiment, participants were seated in front of a computer where the experiment was running with MFC Praat (Boersma & Weenink, 2016). In each gate, the procedure was similar to that in section 3.2.3. First, participants read the instructions that appeared on the computer screen and, once they were ready, they pressed OK on the screen to continue. After 2.0 seconds the audio fragment was played. While the audio was played, the screen was empty. 0.5 seconds after the offset of the audio stimulus, two sentence continuations appeared on the screen, one on the left, the other on the right. Participants were asked to listen to each audio fragment in that gate and then complete the sentence selecting one of the two continuations. After choosing one option on the screen, participants clicked the OK button and 2.0 seconds after clicking OK, the next audio stimulus was played. The order of the two sentence continuations on the screen was counterbalanced to avoid any left/right preference among participants.

The stimuli in each gate were randomized for each participant to avoid a sequence effect. Participants were not forced to make a choice under time pressure. The audio fragments were presented in three consecutive gates, from gate *a* to gate *d*, to make sure that the information participants heard was incremental.

3.3.4 Results

We obtained a total of 4320 responses (3 gates × 40 stimuli × 36 participants). In general, participants were successful in correctly deciding which of the two clause types were intended by the speaker, as illustrated in Figure 5. In gate *a*, where participants only listened to the subject, participants chose a question continuation 54.6% of the time when it was originated from questions. In other words, their overall response accuracy was 54.6% when the intended clause type was question; when the intended clause type is declarative, their overall response accuracy was 59%. The chi-square analysis showed that there was a significant association between the clause type intended by the speaker and the participants' responses, $x^2 = 26.73$ (1), $p < 0.001$. In gate *b*, where participants listened to both the subject and the adverb, the overall response accuracy was 59.7% when the intended clause type was question, and 64.6% when the intended clause type was declarative. Similarly, the chi-square analysis showed that there was a significant association between the clause type intended by the speaker and the listeners' responses, $x^2 = 85.27$ (1), $p < 0.001$.

In gate *d*, where participants listened to subject, adverb and verb-*le diănr*, the overall response accuracy was 62.1% when the intended clause type was question, and 72.1% when the intended clause type was declarative. The chi-square analysis showed that there was a significant association between the clause type intended by the speaker and participants' responses, $x^2 = 169.80$ (1), $p < 0.001$.

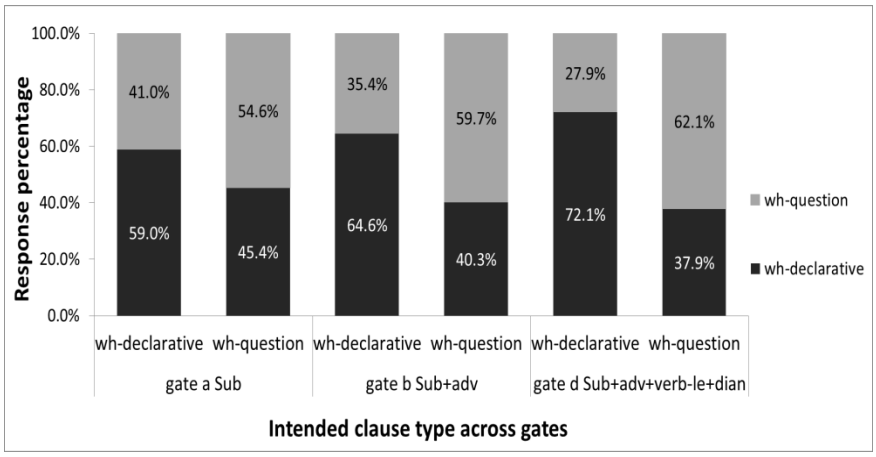

Figure 5. Listener's response in percentage (%) in gate *a*, *b* and *d*

To investigate whether participants' responses can be predicted on the basis of the intended clause type by the speaker in each gate, we also ran a mixed effects logistic regression using the lme4 package in R, with the intended clause type (declarative or a question) as an independent variable, participants' responses (declarative or question) as a dependent variable and items and participants as random factors. Specially, we first ran a null model with participants' responses as a dependent variable, and participants and items as random factors. A second model included in addition the intended clause type as a fixed effect factor to see whether the model was improved. Finally, we ran a third model that included participants' responses as a dependent variable, the intended clause type as a fixed-effect factor, and participants and items as random factors, allowing by-participant and by-item random intercepts, and by-participant and by-item random slopes for the intended clause type. Model fit was compared using the likelihood ratio test (Pinheiro & Bates, 2000; Bolker, Brooks, Clark, Geange, Poulsen, Stevens, & White, 2009). See Appendix B for the details of the fitting model in each gate. The detailed mixed effects results are summarized in Table 3. In general, participants' response on the clause type in each gate can be predicted on the basis of the intended clause type by the speaker.

Table 3. Summary of the results of the mixed effects logistic regression between participants' responses and the intended clause type in each gate.

	Estimate β	Std. Error	z-value	p-value
Gate *a*	0.62	0.30	2.10	< 0.05
Gate *b*	1.51	0.29	4.00	< 0.001
Gate *d*	1.58	0.17	9.35	< 0.001

3.3.5 Discussion

The audio-gating experiment aims at investigating at which point the prosodic differences between *wh*-questions and *wh*-declaratives can be perceived and utilized to predict the sentence clause type. The results from gate *a*, *b* and *d* demonstrate that listeners can make use of prosody to anticipate clause types, distinguishing *wh*-questions and *wh*-declaratives before reaching the *wh*-word *shénme*. In gate *a*, where only the subject of the sentence is heard, there is already a preference for the correct clause type. As the gate goes from *a* to *b* and then to *d*, the more information is perceived by participants, the more accurate their responses are.

With respect to the clause type anticipation, when a sentence unfolds over time and listeners can interpret the information and formulate a prognosis of which clause type the sentence may further unfold, we can hence claim that they can anticipate the clause type. By adopting the audio-gating paradigm, we distinctly collected each prognosis of the clause type in different sentence fragment lengths. The results of the audio-gating study demonstrate that, although *wh*-questions and *wh*-declaratives are string identical, listeners can anticipate their clause types as early as the beginning of the sentences by utilizing the available prosodic cues.

3.4 General discussion and conclusion

The aim of this study was to examine whether and when Mandarin listeners can anticipate the clause type based on pure prosodic cues, by conducting a case study on *wh*-questions and *wh*-declaratives. To achieve this, we first conducted a perception experiment on the identical string of *wh*-questions and *wh*-declaratives to investigate whether listeners can identify the clause type intended by the speaker by perceiving the prosodic differences. The results of the perception experiment served as an experimental baseline for clause type anticipation, since they demonstrated that listeners can indicate the clause type correctly by hearing the prosodic cues of the clause type.

The audio-gating experiment directly addresses the questions of whether listeners can anticipate clause types and when exactly the clause type anticipation takes place based on prosody, by testing different lengths of audio fragments of *wh*-questions and *wh*-declaratives. We obtained listeners' prognosis of the clause type in all the audio-gates and found that listeners can make use of the (limited) prosodic cues to predict the clause type, already from the subject of the sentence. It should be noted that we didn't provide the listeners any prior contexts other than the fragments of *wh*-questions and *wh*-declaratives; in other words, prosodic information is able to serve as the only cue for clause type anticipation in a sentence appearing out of the blue. Further, the more prosodic information is used by listeners, the more accurate their anticipation on clause type is.

With respect to clause type and prosody, our findings also support that, if clause types are marked differently in their specific prosody in a language, these specific prosodic markings (i.e. pitch and duration in our study) are stored in auditory memory and actively used as criteria/filters for listeners' identification of clause

types (Gerard & Clement, 1998). Lastly, concerning anticipation, our findings also imply that listeners seem to make an online assessment of the limited prosodic cues they hear with a reference to the common prosodic marking in each clause type stored in memory and anticipate the most plausible clause type or interpretation accordingly based on the existing prosody.

Although the audio-perception and gating studies can help us to determine whether and at which point prosody is utilized by listeners to identify and anticipate clause types, they are offline studies that cannot provide evidence for the role of prosody on clausal typing during online processing. This is the limitation of the gating paradigm. Different from the behavioral studies like audio-gating, event-related potentials (ERPs) provide a continuous measure of auditory processing with an excellent temporal resolution, and hence serve as an ideal measure for revealing the role of prosody in clausal typing during online processing. We continue this topic of clausal typing by reporting 2 ERP studies in Chapter 6. In the next Chapter, we will first discuss the licensing of *wh*-existentials as the evidence collected in the audio-gating study can illuminate the discussion of the licensing of *wh*-existentials to some extent.