Clause Type Anticipation based on the Prosody of Beijing Mandarin? Evidence from Taiwan Mandarin listeners

Wang, Chih Chieh

Clause Type Anticipation based on the Prosody of Beijing Mandarin? Evidence from Taiwan Mandarin listeners

by

Wang, Chih Chieh 王致婕

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Arts in Linguistics

Faculty of Humanities, Leiden University

July 31st, 2023

Acknowledgements

First of all, I thank my supervisors, Dr. Leticia Pablos Robles and Prof. Lisa Cheng, for their invaluable guidance and encouragement throughout my degree. Their constant support and constructive feedback have greatly assisted me in completing this thesis.

I am grateful to my close friends who have been by my side during difficult moments in the past months, and to my friends in Taiwan who participated in and helped promote the experiment, allowing me to eventually collect all the data in time.

Last, I deeply appreciate my parents and my sister for their unconditional support and understanding.

This thesis is dedicated to my mom and dad.

Table of Contents

Abstract

The current study offers the first clause type anticipation based on the prosody of Beijing Mandarin for native Taiwan Mandarin listeners, building upon the research of Gryllia et al. (2020). By conducting an online audio-gating experiment, our findings indicated that TM listeners encounter difficulties anticipating *wh*-questions based on BM prosodic cues, whereas the anticipation of declaratives was more successful. The effect of the clause types intended by the BM speaker was significant from Gate II, and the effect of tones in Gate III suggested an intertwined interaction between tone and intonation. Additionally, the impact of the gate and its interaction with the intended clause type was also shown to influence the clause type anticipation.

1. Introduction

1.1 Anticipation

Merely a few decades ago, language processing was primarily supported by the integration approach, which explained how people comprehend incoming information with previous knowledge and establish novel representations (Ferreira & Chantavarin, 2018; Long & Lea, 2005). In recent years, psycholinguistics has turned its focus on prediction, a mechanism that facilitates comprehension by anticipating upcoming information (e.g., Bansal et al., 2018; Clark, 2013).

The human brain is built for anticipating upcoming events, externally from daily activities such as biking in the hectic streets knowing when to turn or to break, to internally thinking of the potential scenarios to be prepared in advance. In a natural conversation, there is often a seamless transition with minimal gaps or overlaps between speakers (Kuperberg & Jaeger, 2016). We seem to know when the interlocutor is about to end, then take turns and keep up the flow of a conversation. Compelling evidence has been discovered that these predictions of rapid exchanges in conversations are based on the lexical and syntactic information that listeners have heard (e.g., de Ruiter et al., 2006; Garrod & Pickering, 2015; Magyari & de Ruiter, 2012).

To achieve effective and accurate information during language processing, whether human or artificial beings, cognitive systems heavily rely on information from the past to expect forthcoming events (Ferreira & Qiu, 2021). Even though some researchers have argued that predicting upcoming information may take greater processing resources (e.g., Jackendoff, 2002; Van Petten & Luka, 2012), a growing body of research have examined the importance of prediction in effective language comprehension.

In addition, many of them have emphasized lexico-semantic processing. As human beings are meaning-oriented creatures, previous findings show that participants are able to predict the next words based on semantic constraints. For instance, *the day was breezy so the boy went outside to fly a* __, often, comprehenders will anticipate the word *kite* to fill in the blank (example from Ferreira & Qiu, 2021). Early eye-tracking studies (e.g., Balota, et al., 1985; Ehrlich & Rayner, 1981) demonstrate that participants fixate less on predictable words than on less predictable ones. Moreover, evidence of the N400 effect, an Event-Related Potential (ERP) component, was observed in semantic violations and suggested an inverse function of expectancy words from the subjects' cloze probability (e.g., Kutas & Hillyard, 1984).

Although prediction on syntactic categories seems to have received less attention, the logic behind anticipating a specific word based on semantic constraints inherently implies that syntactic anticipation is also involved (Ferreira & Qiu, 2021). From the above example, besides the word *kite* was predicted, a singular noun was anticipated at the same time.

Interestingly, in the realm of human communication, it is not just the words we choose that shape our conversations, but also how we deliver them. The way we speak, with its unique acoustic cues of pitch, tempo, duration, and amplitude, plays a fundamental role in our daily interactions. These acoustic elements, collectively known as prosody, are crucial in conveying not only the emotional state, identity, and attitude of a speaker but also providing both lexical and post-lexical information (Mennen & de Leeuw, 2014).

1.2 Prosody

Prosody encompasses the suprasegmental features of language and is present in every utterance and across all languages (Cutler et al., 1997). For example, if the phrase *Mary bought an appl*e is said with a rising pitch, it often represents a question, while the same word string said with a falling pitch is mostly considered a statement. Furthermore, in many languages like English, different stress patterns in words carry various linguistic functions. For instance, when the stress is on the second syllable of *permit* /pɚˈmɪt/, it represents a verb meaning giving authorization or consent to someone to do something, On the other hand, if the stress is on the first syllable, /ˈpɝmɪt/, it is used as a noun to refer to an official document giving authorization.

Understanding the significance of prosody in communication opens a fascinating door into the complexities of human language processing and comprehension. It allows us to explore how the manipulation of pitch, duration, and other acoustic cues can influence the meaning and interpretation of spoken language. Since anticipation is a mechanism of preparedness in language processing (Ferreira & Chantavarin, 2018), can there be anticipation based on prosody?

1.2.1 Prosody Anticipation

The study of anticipation is fundamental to understand language processing. An online discourse comprehension eye-tracking task (Ito & Speer, 2008) examined the role of pitch accents in American English. Their findings indicate that listeners are capable of using contrastive accents to predict the reference of noun phrases. Another study (Nakamura et al., 2012) investigated the effect of intonation contrasts on relative clause to predict a syntactic structure in Japanese. In addition to prosody influencing syntactic analysis and costing at disambiguating information, the results of the study also offer initial evidence supporting the prediction of pre-head structures based on prosodic and contextual information in a head-final construction.

To some degree, prosody enhances the anticipatory of comprehension. In Stromswold et al. (2002)'s eye-tracking experiment, participants listened to an active or passive sentence while seeing two images on the screen. They found that the vowel duration of the verbs helps anticipate the agents of the sentences in English (e.g., the /u/ sound in *the girl was pushing the boy* and t*he girl was pushed by the boy*). Another study (Roll et al., 2010) of ERP online processing of Central Swedish discovered that high stem tones enable listeners to activate their associated suffixes (single or plural). In other words, the tone-inducing suffixes, which are assigned morphophonologically, are predicted by the high accent tones.

Furthermore, cross-linguistic study (Choi et al., 2016) presented that prosody in first language (L1) could affect second language (L2) acquisition. In their Cantonese-English 1-year-long research, they found a direct prosodic transfer from Cantonese lexical tones to English lexical stress, which facilitates native Cantonese children from early-year primary school predicting in English reading comprehension.

Although studies on anticipation across languages provide insights into variation and universality, the prosody anticipation on tone languages, to date, poses a significant challenge of anticipation in language processing. Specifically, tone languages depict a unique way of utilizing prosody to resolve linguistic ambiguity and indicate sentence types (Hsu & Xu, 2020).

1.3 Mandarin Chinese

Mandarin Chinese is one of the widely spoken language among tone languages which there are more than 1 billion speakers worldwide (Ethnologue, 2023). Tone is a prosodic feature that signals the pitch patterns in lexical or morphological contrasts, whereas intonation indicates the melodic aspect (Xu, 2019). Speakers of tone languages primarily use pitch variation to convey both lexical and post-lexical information in a nuanced and intricate manner, which differs from speakers of non-tone languages (i.e., most Indo-European languages) who mainly use pitch variation for post-lexical information (Zou et al., 2015).

There are four lexical tones (henceforth referred to as T1-T4) and a neutral tone (referred to as T0) in Mandarin. Namely, T1 is a high-level tone, T2 is a mid-rising tone, T3 is a low-dipping tone, T4 is a high-falling tone, and T0 is an unstressed short syllable at a non-initial position in a word or phrase (Yang, 2016). However, the low-dipping T3 only exhibits in isolation or at the boundary position; in the non-utterance final position, it occurs as a low tone without the final rise (Yang, 2016; Xu, 2004; Wang et al. 2020).

Additionally, both phrase-level and utterance-level prosody play a crucial role in shaping the language (Yang, 2016). Encoding F0 (fundamental frequency) variations in tone languages can be challenging. As F0 differences serves multilevel functions, such as lexical tones and intonation (Chen, 2022), it can lead to either congruent or incongruent cues. For instance, several studies (e.g., Liu et al., 2016; Yuan, 2006) have investigated whether the presence of T2 or T4 at the utterance-final position of a question or statement affects the identification of sentence types, or whether intonation influences the identification of tones (Gryllia et al., 2020).

1.3.1 Tone and Intonation

Regarding the intricate interplay of tone and intonation, question intonation in Mandarin is frequently discussed (Yang, 2016). For example, in Yuan's (2011) perception study, the results showed that for native Mandarin listeners from northern China, question intonation was more difficult to identify with a rising tone (T2) in final position, but easier with a falling tone (T4).

Recent research (Liu et al., 2022) found that, in a semantically neutral context, Mandarin listeners from northern China (Mandarin dialect area) tone identification was generally better than intonation. In contrast, identifying intonation of sentence types was influenced by the final lexical tone, the rising tone (T2) and falling tone (T4) were equally difficult in questions. Furthermore, their results demonstrated that semantically constrained context facilitates participants distinguishing tone and intonation information, particularly sentences with T4 in the final position.

The captivating relationship between tone and intonation in Mandarin continues to pique the interest of researchers. Khlifat (2016) examined speeches of three native Mandarin speakers, focusing on tones T2 and T4 in questions and statements. The study concluded that the F0 at the end of a yes/no question tends to be higher than in a statement and argued that there was a simultaneous tonal and intonational contours on a single word.

In addition, Zeng et al. (2004) analyzed the interaction between tone and three modalities: declaratives, yes/no questions (with interrogative marker *ma*), and rhetoric questions (without *ma*) of five speakers from northern and southern China. They examined the F0 and intonation, and the results indicate that tonal realizations were significantly influenced by the utterance types. For instance, T1 became ascending in rhetoric questions interrogative sentences; the duration of T1, T2, and T3 in sentence-final position was the same in declaratives and rhetorical questions but much shorter in yes/no questions.

However, many previous studies in the speech production and perception literature that focus on Mandarin use the term in a broad sense, without specifying which regional variety of Mandarin was investigated (Chang, 2010). It is important to note that different varieties within a language can manifest and convey information in distinct ways. Similar to individuals from England and America being able to communicate effectively, the English language is utilized in diverse manners depending on the specific context and region.

1.3.2 Beijing Mandarin and Taiwan Mandarin

Although Mandarin Chinese has been spoken in China for a considerable period of time, it has only recently attained the status of a standard language. Due to the social and political changes, Beijing Mandarin was recognized as the standard pronunciation of Standard Chinese (i.e., *pǔtōnghuà* ‘common speech’) in the People’s Republic of China (PRC) in 1956 (Chang, 2010). Likewise, Taiwan Mandarin (i.e., *guóyǔ* ‘national language’) was implemented in formal sectors and became the dominant language of the island after 1949. While the two varieties of Mandarin hold equal prestige and share many similarities, they also possess unique characteristics (Torgerson, 2005).

In recent years, acoustic studies on two major varieties of Mandarin Chinese have consistently demonstrated distinct differences between Beijing Mandarin (BM) and Taiwan Mandarin (TM). For example, when comparing the realization of tones in BM and TM, it was found that TM exhibits a slower tempo, a lower register, and a different realization of T3 (Chang, 2010). In Chang (2010)’s tone-gating perception experiment of these two regional varieties on native and nonnative listeners, they found that BM and American English speakers who learned BM as a L2 both had difficulty identifying T3 in isolation in TM which was misidentified as T4.

Whereas TM listeners had significantly lower accuracy for recognizing intact T4 in BM which was perceived as T3.

When it comes to intonation, Tseng (2004) analyzed the speech production from radio announcers of both varieties and concluded that BM seemed like a stress-timed language that demonstrated by a more significant pitch fall and greater contrast in duration and intensity, while TM was more of a syllable-timed that commonly perceived as sounding softer and gentler. Specifically in declarative sentence, TM had a slower tempo. Whereas BM had a higher F0 which also corresponded that the speakers of BM frequently perceived speaking in a higher voice. Moreover, F0 manipulation seemed freer in BM, where downdrift was used in TM instead.

Correspond to Tseng (2004)'s conclusions, another speech production study (Hsu & Xu, 2020) focused on sentence-final particles (SFPs) and *wh*-words that carried interrogative or indefinite readings among three sentence types: *wh*-questions, yes/no questions, and statements found that the two varieties represented syntactic-semantic information differs prosodically while maintaining lexical tones. For example, BM speakers marked distinctly three sentence types with F0 contours, while TM speakers mainly differentiated statements from questions.

Torgerson (2005) examined the differences in pitch and tone register between BM and TM speakers in spontaneous interview, spontaneous descriptive, and controlled read sentential speech. The findings suggest that TM has a slightly lower tone register in comparison to BM, and interestingly, the speech style was not a significant predictor.

To enhance our comprehension of Mandarin, it is essential to consider the acoustic distinctions between its two main varieties. Conclusions derived from previous studies that primarily focus on Mandarin features in BM may not be applicable to TM, and vice versa. While production studies offer valuable insights into the prosodic characteristics of these varieties, conducting perception studies with BM listeners and comparing them to TM, or conversely, can both further enhance our understanding of Mandarin processing.

Next, we'll first look into the syntactic features of Mandarin and whether prosody function as a mean to type a clause in the language. Then, our attention will shift to prosody anticipation in Mandarin, an area that remains rather underexplored in language processing.

1.4 Clausal Typing in Mandarin

Mandarin is a *wh*-in-situ language, in other words, *wh*-words are presented at the base position in a clause (e.g., Mary bought *what*?). Whereas many Indo-European languages, such as English or Dutch, are *wh*-movement languages. In other words, these languages tend to position *wh*-words in the beginning of a sentence (e.g., *What* did Mary buy?). However, typing a *wh*-question in Mandarin seems to be more complex compared to *wh*-movement languages, where the *wh*-word is fronted to type the clause, or other *wh*-in-situ languages like Japanese, utilize a *wh*-particle (e.g., *ka*) at the end of a clause to type a *wh*-question. Although Mandarin has particles like *ne* (also known as a SFP) that can function as a *wh*-particle, its usage is optional in *wh*-questions and can also be employed in yes/no questions (Yang, 2018).

Yang (2018) conducted thorough study on the clausal typing mechanism in BM. Their production and perception experiments demonstrated that prosody serves as a clausal typer to distinguish between *wh*-questions and *wh*-declaratives, where the latter consists of *wh*-words like *shénme* 'what' carrying non-interrogative meanings. Moreover, they presented electrophysiological evidence from two auditory ERP experiments to support the effects of prosody on clausal typing. Based on their findings, they proposed an extended clausal typing hypothesis, suggesting that in the case of *wh*-questions, *wh*-in-situ languages like Mandarin can use prosody to type the clause.

1.5 Clause type anticipation in Mandarin: Gryllia et al. (2020)

Further develop from Yang's (2018) research, previous study conducted by Gryllia et al. (2020) examined the anticipation of prosody of clause types between the declarative and *wh*-question in BM. Since these two clause types only differ in the *wh*-word *shénme* 'what' and its non-*wh*-counterpart (e.g., Mary bought *what*? vs. Mary bought *apples*.), this provides an opportunity to investigate the role of prosody in both the pre-wh-word and non-*wh*-counterpart regions, as well as to examine whether these prosodic features are utilized in anticipation.

They first carried out a production experiment where they found that native speakers of Beijing Mandarin employed prosodic cues to distinguish the two clause types without relying on explicit morphosyntactic markers. Additionally, they observed that the two clause types exhibited differentiation from the beginning, manifesting differences in duration, F0, and intensity. Specifically, *wh*-questions were produced with a faster speech rate, higher F0, a narrower F0 range in verbs of T1 or T4, and a higher intensity range on the second syllable of the adverb.

Subsequently, a perception experiment was conducted to determine if native listeners could utilize these prosodic cues to anticipate the two clause types. Overall, the participants' anticipations were significantly associated with the intended clause types of the speakers. For example, in the first gate (referred to as gate-a in their study), participants heard two prosodic cues in the audio fragments before anticipating the following sentence continuations. Firstly, there was a higher F0 maximum (H) on S2 (i.e., the second syllable), and secondly, the duration of S1 and S2 was shorter in the *wh*-questions. Similarly, in the second gate (gate-b), *wh*-questions exhibited a higher F0 maximum (H) on S4, along with shorter durations of S3 and S4. As a result, the prosodic cues between the first and second gates were similar, as indicated by the participants' responses, and did not demonstrate a significant difference. However, in the last gate (gate-c), the tones of the verbs played a crucial role in the observed results. Especially for the case of T3, listeners confused *wh*-questions with declaratives which suggests that the low-dipping tone affects other prosodic cues they received in the previous gates.

1.6 The current project

The present study builds upon the previous research (Gryllia et al., 2020) by applying the audio-gating experiment with identical stimuli to TM listeners. The objective is to explore the nature of clause type anticipation based on prosodic cues and address the gap in prosody perception between BM and TM. By considering the prosodic differences highlighted in the aforementioned literature, our aim is to discover whether the findings of Gryllia et al. (2020) can be replicated.

Our primary hypothesis is that although TM listeners may initially struggle to anticipate the upcoming clause types from the first gate, as more information is gradually revealed, they are expected to be able to anticipate both declarative and *wh*-question clause types. Additionally, we hypothesize that TM listeners will perform better in declaratives compared to *wh*-questions, which aligns with Gryllia et al.'s (2020) findings. Lastly, given the confusion in identifying BM T4 into T3 for TM listeners in Chang (2010) and the tone effect observed in Gryllia et al. (2020), we propose that the tone of the verbs in Gate III will influence anticipation.

To conclude, our research questions are:

1) Can TM listeners anticipate clause types based on BM prosody? If so, from which gate?
2) Is the anticipation in declaratives more predictable than *wh*-questions among all gates?
3) Does the tone of the verbs affect TM listeners' anticipation of the clause types in Gate III?

2. Methods

2.1 Participants

Thirty-seven native speakers of Taiwan Mandarin participated in the online experiment (12 males, 25 females, age *Mean ± SD = 28.49 ± 3)*. They were all born and raised in Taiwan, where Taiwan Mandarin is their dominant language in daily life. Participants were recruited from the experimenter's personal network and did not receive any reimbursement for their participation. All participants reported no language disorders (e.g., dyslexia) or hearing difficulties and provided their consent to the experiment.

Participants who did not complete the experiment were manually or automatically excluded, and the IP location of their device was exclusively limited to Taiwan. Initially, thirty-nine participants' data was collected. The data of two participants were excluded due to either incompletion of the questionnaire or reported having language disorders and hearing difficulties.

2.2 Audio Stimuli

A total of 40 stimuli, consisting of two clause types (*wh*-question and declarative), were used in the experiment. The stimuli were originally employed in the perception experiment conducted by Gryllia et al. (2020). They were selected from one of the participants of the production experiment in the same research, who was a 20-year-old female native speaker of Beijing Mandarin. All the stimuli were converted from .wav to .mp3 files to accommodate the online experiment design, and the volume was uniformly adjusted to 89 dB using MP3 Gain Express v2.5 (https://projects.sappharad.com/mp3gain/).

2.3 Audio-Gating paradigm

Each audio stimulus was divided into 3 fragments (40 stimuli x 3 fragments), a total of 120, using the audio-gating paradigm, manipulating the word and prosodic boundaries. These fragments were played in sequence during Gate I, Gate II, and Gate III. Within each gate, information about the sentence was gradually revealed. The main objective of this experiment was to determine whether prosodic information has a significant impact on the anticipation of the clause types for native Taiwan Mandarin listeners.

Each audio fragment was presented in each gate respectively: Gate I included the Subject in proper names (e.g., *Táo Wēi*), Gate II consisted of the Subject and the Adverb indicating time (e.g., *zuótiān* 'yesterday'), and Gate III consisted of the Subject, the Adverb, the Verb (e.g., *ná* 'bring') plus the perfective marker *le*. All lexical tones except the Verbs were kept consistent among the two clause types (see examples (1)-(3) with the number of syllables and the corresponding tones). In addition, four tones of the Verbs were equally distributed (i.e., 10 each per tone) in Gate III.

(1)

Gate I	
Subject	
S1	S2
[T2	T1]

(2)

Gate II			
Subject		Adverb	
S1	S2	S3	S4
[T2	T1	T2	T1]

(3)

Gate III					
Subject		Adverb		Verb	PERF
S1	S2	S3	S4	S5	S6
[T2	T1	T2	T1	T1, T2, T3, T4	T0]

2.4 Response Stimuli

For each gate, response stimuli of the two clause types were presented counterbalanced on the screen using traditional Chinese characters, after the audio fragment was played. It should be noted that traditional Chinese characters are the script used in Taiwan, whereas simplified characters are used in China (see Appendix A for the two scripts). Both *wh*-questions and declarative sentences as continuations only differed at the *wh*-word and its non-*wh*-counterpart (see Gate I examples in (4) and (5)).

(4)

Wh-question continuation

zuótiān	*tōule*	***shénme***	*gěi*	*báibīng*
yesterday	steal.PERF	**what**	for	Baibing

'stole what for Baibing yesterday?'

(5)

Declarative continuation

zuótiān	*tōule*	***nízi***	*gěi*	*báibīng*
yesterday	steal.PERF	**woolen**	for	Baibing

'stole woolen for Baibing yesterday.'

2.5 Instruction

All instruction was adapted to traditional Chinese character script from the original simplified one in Gryllia et al.'s (2020). Except the word *lùyīn* 'recording' was replaced by *yīndǎng* because *lùyīn* used as 'to record' in TM. A similar text format was used for all gates, with one key difference: Gate I explicitly informed participants that there would be one word which was a proper name (since the two-character name is less common in Taiwan), and Gate II underlined that there would be two words in the audio fragments. For instance, in Gate I, "*This is an experiment in which you listen to a recording to complete a sentence. The recording is a sentence fragment with only one word (person's name). Since the sentence is incomplete, based on the information you hear, choose how the sentence should be said further to be complete. There are two options, one* *declarative and one interrogative sentence, click on the option you think is correct. When you have completed a question, click "OK" to continue.*" (See Appendix A for the full instruction text).

2.6 Procedure

The experiment was conducted online using the Gorilla Experiment Builder (Anwyl-Irvine et al., 2019). It consisted of two phases (Practice phase and Test phase: Gate I-III) and one background questionnaire (see Appendix A for the details). There was no time limit for providing the responses, participants took approximately 16 minutes to complete the whole experiment. They could use any device (e.g., computers, tablets, or mobile phones) of their own choice with earphones or headphones to conduct the tasks.

At the beginning of the experiment, the Participant Information Form and Consent Form from Leiden University Centre for Linguistics (LUCL), which provided details about the anonymity of their data, their right to withdraw from the experiment, and the experiment's procedure, were presented to the participants. Participants were advised to find a quiet environment and turn off notifications on their devices to minimize disruptions during the experiment.

A browser sound check was implemented to ensure the quality of the audio stimuli. Initially, a short music fragment was played, allowing participants to check that the auto-play function of their devices was functioning correctly. In case of any issues, instructions were provided on adjusting the sound settings in their browsers.

After successfully completing the browser sound check, participants proceeded to the Practice phase with the information that the experiment would advance progressively (i.e., in the first part they would only hear one word, in the second part two words…etc.) It consisted of 6 trials (2 trials x 3 gates). First, the instruction slide was presented, and each audio fragment was played every 1.0 seconds after clicking the "OK" button. The screen was empty when the audio fragment was played and 0.3 seconds after the fragment's offset, participants had to select a sentence continuation (either a *wh*-question or a declarative) that was shown on the screen based on what they had heard and clicked "OK" button to proceed to the next trial. There was no feedback provided.

The purpose of the Practice phase was to familiarize participants with the upcoming task. In order to minimize the impact of familiarity, the stimuli used in the practice trials were recorded by the experimenter who was a 27-year-old female native speaker of Taiwan Mandarin. The two sentences were designed and divided into three fragments in a manner that ensured no overlaps with the test stimuli. Participants gradually received information through three gates, and it took average 1:22 minutes to complete the task.

In the Test phase, participants followed a similar procedure as in the Practice phase. They were presented with an additional information slide as a reminder before the instruction slide. This slide indicated that even though certain words from the response stimuli were from Beijing Mandarin (e.g., *tízi* 'grape' which in Taiwan Mandarin is *pútáo*), these regional differences should not influence their answers. Once the audio fragment was played, two sentence continuations were displayed on the screen, and participants had to select one based on what they had heard. They could proceed to the next trial by clicking the "OK" button.

All trials within each gate were randomized. The sentence continuations displayed on the screen were counterbalanced to ensure an equal presentation of the two options and to avoid the left or right preference of the participants. The audio fragments were randomized and played consecutively, starting from Gate I to Gate II, and then Gate III. This sequential presentation allowed participants gradually receive accumulated information throughout the Test phase. On average, it took 11 minutes and 37 seconds to complete the three gates.

After completing the gating experiment, participants were directed to a questionnaire where they provided additional information about their demographics and language background (see Appendix A section 2). It took about 2 minutes to complete. Participants were required to self-report any hearing problems or medical diagnoses of language disorders. They were then asked to rate their Chinese proficiency in reading and listening comprehension using a 7-point Likert scale (1 = very poor, 7 =excellent). A 5-point Likert scale ranging from “strongly disagree” to “strongly agree” was used to assess whether Mandarin Chinese was their primary language in their living environment and the most frequently contact language on a daily basis.

Additionally, participants were asked to indicate whether they acquired another native language (e.g., Taiwanese Min, Hakka, or other languages) and rate their proficiency in that language, if applicable. They were also requested to provide the percentage (which should add up to 100%) of their usage for each language in day-to-day life. Furthermore, participants were given the opportunity to provide voluntary feedback or any other advice. On the last page of the questionnaire, the experimenter expressed gratitude to the participants for their participation and instructed them to press “continue” bottom to finish the online experiment.

3. Statistical analysis

We analyzed the data employing generalized linear models (e.g., mixed-effects logistic regression) with binary dependent variable (*wh*-question and declarative) to investigate the likelihood of participants’ responses based on the clause type that was intended by the speaker. All the subsequent analyses using the *lme4* Package (Bates et al., 2015) in R, specifically, using anova function to compare the fitted models and chi-square tests to examine the association between dependent and independent variables.

To examine the effects within each gate, we initially ran a null model (m0) with participants' responses as the dependent variables and participants and items as random intercepts for each gate. We then introduced a second model (m1) by including the intended clause type as a fixed effect to assess whether it improved the null model. For Gate I, the null model was retained, while for Gate II and Gate III, the second model demonstrated a better fit. We conducted two additional models for Gate III: one included the Verb tone as a second fixed effect (m2), and another involved adding the interaction between the clause type and Verb tone (m3). The third model (GateIII.m2) with the Verb tone as the second predictor yielded the best fit. Then, we also investigated the effect of Verb tone solely in Gate III (Tone.m1) compared to the null model.

To determine whether there was a significant association between participants' responses and the intended clause type in each gate, chi-squared analyses were performed. Moreover, the distribution of participants' responses regarding the intended clause type was presented in percentages. Additionally, we also examine the association between the response and clause type for each tone in Gate III.

Lastly, considering that overall performance slightly improved as more information was released in the audio stimuli from Gate I to Gate III, we conducted additional analyses to examine the effect of the gate. We ran a null model (allGate.m0) with participants' responses as the outcome and participants and items as random factors. Subsequently, a second model (allGate.m1) included the intended clause type as the first predictor, a third model (allGate.m2) added the gate as the second predictor, and finally, a fourth model (allGate.m3) incorporated the interaction between the intended clause type and the gate. The last model was determined to be the best-fitting model. To validate, we also ran a model with the effect of gate alone (Gate.m1) in comparison to the null model; (see all the best-fitting models in Appendix B).

4. Results

In total, we received 4440 responses (3 gates x 40 stimuli x 37 participants). However, there were 20 responses that participants did not choose any of the continuations and pressed “continue” to skip to the next question. The total number of sentence continuations obtained and analyzed in the subsequent sections was 4420. Among these, Gate I had 1470, Gate II had 1477, and Gate III had 1473 (T1: 368, T2: 369, T3: 369, T4: 367).

4.1 Effect of intended clause type in each gate

Overall, the intended clause type from the Beijing Mandarin speaker in the audio stimuli significantly affected the responses of Taiwanese listeners from Gate II; see **Table 1**.

For Gate I, the null model with random effects was retained in comparison with the one that included the intended clause type as a fixed effect. Hence, there was no significant effect of the intended clause type on the participants’ responses. However, for Gate II, the model improved when added the intended clause type [X^2 (1) = 37.15, $p < .001$]. Similarly, the model improved when adding the intended clause type in Gate III [X^2 (1) = 42.65, $p < .001$].

Table 1

Summary of the results of the mixed-effect logistic regression within each gate

	Estimate β	***Std. Error***	***z*-value**	***p*-value**
Gate I (Intercept)	-0.20	0.15	-1.27	= .202
Gate II	0.71	0.12	6.03	< .001**
Gate III	0.73	0.11	6.45	< .001**

Note. The asterisk indicates the significance (< .05*, < .001**).

4.1.1 The case of Gate III

Furthermore, a Bartlett test showed the variances between tone distributions were not significantly unequal [$p > .05$]. When added the Verb tone as a second fixed effect, the model improved [$X^2 (3) = 67.91, p < .001$], but did not improve when the interaction between the clause type and Verb tone was added. We can, therefore, conclude that the effects of intended clause type and Verb tone were significant to listeners' responses; see **Table 2**.

Table 2

glmer results, Response ~ Clause Type + Tone + (1 | Participant) + (1 | Item)

	Estimate β	***Std. Error***	***z*-value**	***p*-value**
Intercept	-0.38	0.17	-2.19	< .05*
clause type Question	0.77	0.12	6.62	< .001**
Tone 2	-0.22	0.16	-1.36	= .17
Tone 3	-1.28	0.17	-7.53	< .001**
Tone 4	-0.41	0.16	-2.60	< .05*

Note. The asterisk indicates the significance (< .05*, < .001**).

4.1.2 Effect of Verb tone in Gate III

In addition, we analyzed solely the effect of Verb tone on responses in Gate III by comparing a null model and another one that added Verb tone as a fixed factor. The second model was a better fit and that the results were similar to the previous best-fitting model; see **Table 3**. In other words, T2, T3 and T4 tend to have more declarative responses than T1, and this tendency is significant when the Verb bore T3 and T4.

Table 3

glmer results, Response ~ Tone + (1 | Participant) + (1 | Item)

	Estimate β	**Std. Error**	***z*-value**	***p*-value**
Intercept	0.008	0.16	0.05	= .96
Tone 2	-0.21	0.15	-1.37	= .17
Tone 3	-1.24	0.17	-7.42	< .001**
Tone 4	-0.40	0.16	-2.56	< .05*

Note. The asterisk indicates the significance (< .05*, < .001**).

4.2 Association between response and intended clause type

Next, we will present participants' responses in percentage on the basis of intended clause type. In general, participants exhibited anticipating the declaratives more easily than *wh*-questions; see **Figure 1**. We also performed a series of chi-squared tests to see whether there was a significant association between the intended clause type and participants' responses.

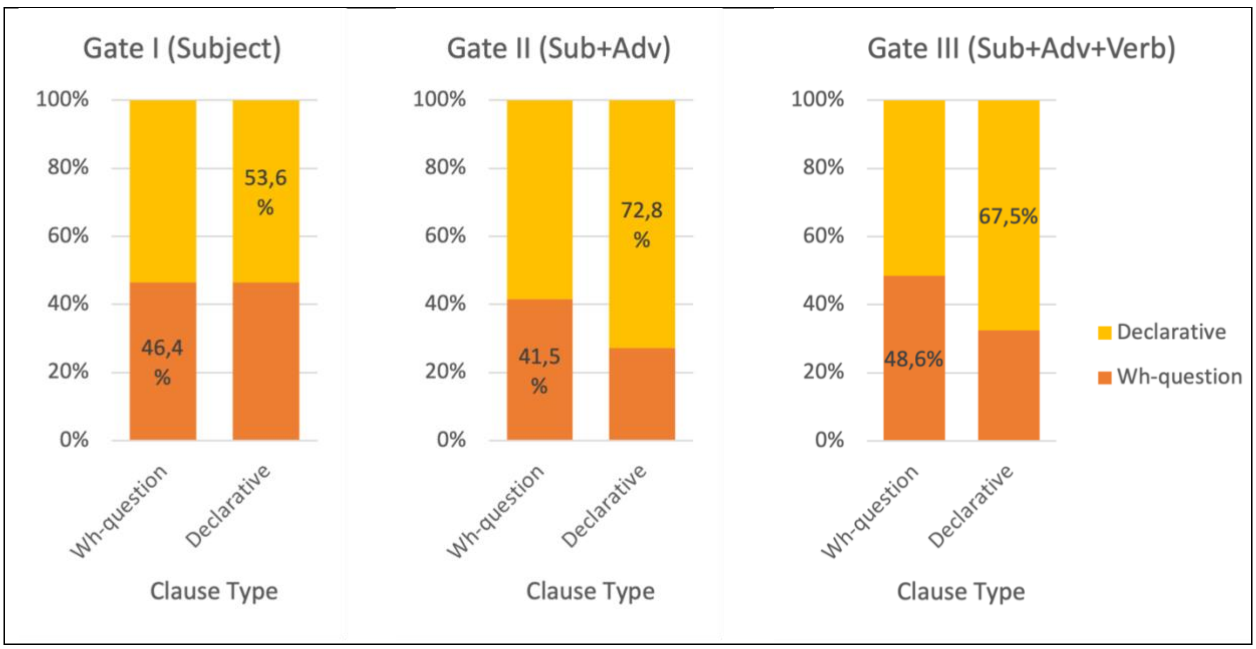
In Gate I, where participants heard a disyllable Subject, they chose a *wh*-question continuation 46.4% of the time when the audio fragment intended from a *wh*-question. When the intended clause type originated from a declarative, listeners chose 53.6% of the time a declarative continuation. A Pearson's chi-squared test showed that there was no significant association between the intended clause type by the BM speaker and the responses from the TM listeners [x^2 (1) = 0, $p = 1$].

In Gate II, the audio fragment consisted of Subject and Adverb, participants performed successfully anticipating the intended declarative clause type 72.8% of the time after hearing the fragment, while only 41.5% of the time they chose a *wh*-question when the intended clause type was a *wh*-question. As opposed to the results of Gate I, a chi-square analysis indicated that there was a significant association between the type of clause that was intended by the speakers and the participants' responses [x^2 (1) = 33.33, $p < .001$].

In Gate III, participants heard the audio fragment with Subject, Adverb, Verb, and the perfective marker *le*. For the fragments that originated from declaratives, listeners chose a declarative continuation more often than a *wh*-question (67.5%). In consistent to the previous gates, less frequent of the time they chose a *wh*-question continuation when the original was a *wh*-question (48.6%). A chi-square analysis also showed that the association between the intended clause type and participants' responses was significant [x^2 (1) = 39.36, $p < .001$].

Figure 1

Participants' responses in percentage (%) to audio stimuli in Gate I, Gate II, and Gate III



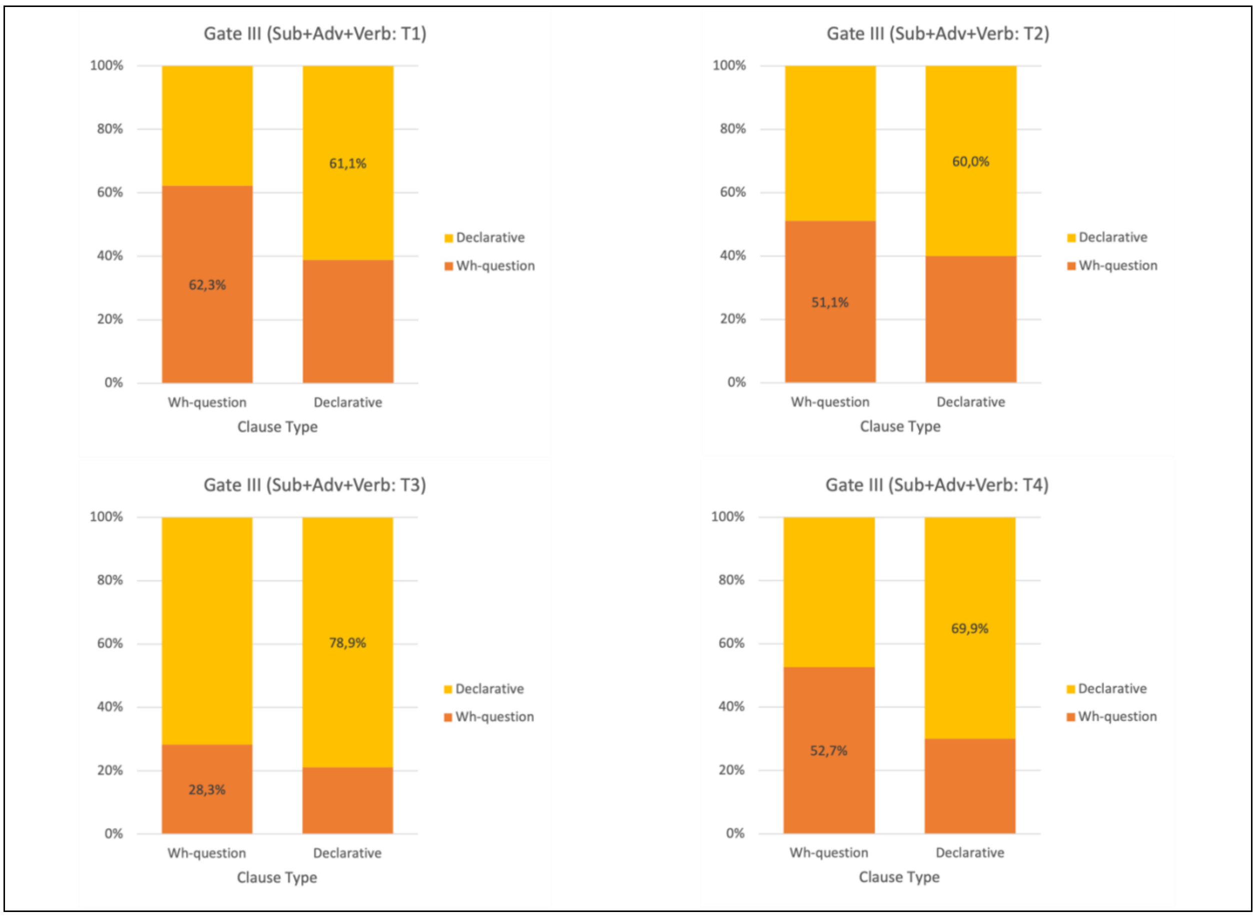
4.2.1 Association within tones in Gate III

From the first analysis of the effects, we knew that Verb tone also played a role in affecting listener's responses, hence, we ran an extra chi-squared test to see whether there was a significant association between the response and Verb tone [x^2 (3) = 57.8, $p < .001$].

The responses of each tone in Gate III were presented respectively in **Figure 2**. As depicted in the figure, participants chose the declaratives more often than the *wh*-questions when the speaker intended declaratives in all four tones (T1:61.1%, T2:60%, T3:78.9%, T4:69.9%). Moreover, participants were able to select *wh*-questions slightly above the chance level when the intended clause type was *wh*-questions in T1 (62.3%), T2 (51.1%), and T4 (52.7%), except for T3 (28.3%). Interestingly, listeners selceted a *wh*-question rather more often while the intended clause type was a *wh*-questions compared to the clause type of declarative when the Verb bore T1.

Figure 2

Participants' responses in percentage (%) to audio stimuli in each tone of the Verb in Gate III



The chi-squared analyses showed that the association was significant when the Verb bore T1 [$x^2(1) = 20.11$, $p < .001$], T2 [$x^2(1) = 4.57$, $p < .05$], and T4 [$x^2(1) = 19.42$, $p < .001$]. On the other hand, there was no significant association between the responses from TM listeners and the clause types intended by the BM speaker when the Verb bore T3 [$x^2(1) = 2.56$, $p = .11$]. In particular, the distribution of T3 differs than the others.

As the percentage of time selecting a sentence continuation reflects the accuracy of the responses, we can see that participants overall performed better in anticipation of the clause types when the Verb tone in T1, T2, and T4 compared to the previous gates; see **Table 4**.

Table 4

The distribution according to the intended clause type within each tone

	Gate I	**Gate II****	**T1****	**T2***	**T3**	**T4****
Wh-question	46.4%	41.5%	62.3%	51.1%	28.3%	52.7%
Declarative	53.6%	72.8%	61.1%	60%	78.9%	69.9%

Note. The asterisk indicates the significance of the association (< .05*, < .001**).

4.3 Effects of gate and intended clause type

Finally, we examined whether the intended clause type and gate affected participants' responses in all gates. To make sure that unbalanced data within gates did not have significant effect, we also ran an additional homogeneity of variances Bartlett test [$p > .05$]. The model improved when we added the intended clause type as a first fixed effect [$X^2\,(1) = 50.25$, $p < .001$]. By adding gate as a second fixed effect, the model improved further [$X^2\,(2) = 48.56$, $p < .001$]. Furthermore, the last model with interaction between intended clause type and gate turned out to be the best-fitting model [$X^2\,(2) = 27.25$, $p < .001$]; see **Table 5**.

Table 5

glmer results, Response ~ Clause Type + Gate + Clause Type:Gate + (1 | Participant) + (1 | Item)

	Estimate *β*	***Std. Error***	***z*-value**	***p*-value**
Intercept	-0.18	0.13	-1.34	= .18
clause type Question	-0.002	0.11	-0.02	= .98
gate II	-0.90	0.11	-7.85	< .001**
gate III	-0.63	0.11	-5.61	< .001**
clause type:gate II	0.69	0.16	4.34	< .001**
clause type:gate III	0.72	0.16	4.62	< .001**

Note. The asterisk indicates the significance (<0.05*, <0.001**).

The above findings were validated as we ran an additional model solely looking into the effect of the gate and random effects, which demonstrated that the listeners' responses were significantly affected by the gates; see **Table 6**.

Table 6

glmer results, Response ~ Gate + (1 | Participant) + (1 | Item)

	Estimate *β*	***Std. Error***	***z*-value**	***p*-value**
Intercept	-0.18	0.12	-1.49	= .13
gate II	-0.54	0.08	-6.90	< .001**
gate III	-0.26	0.08	-3.32	< .001**

Note. The asterisk indicates the significance (<0.05*, <0.001**).

4.4 Questionnaire results

A 5-point Likert scale (1 = strongly disagree, 5 = strongly agree) was used to measure the questions that Mandarin Chinese was the primary language in the living environment and the most contact language in daily life. And a 7-point Likert scale (1 = very bad, 7 = excellent) was employed to assess proficiency in Mandarin Chinese in both listening and reading; see **Table 7**.

Table 7

The results of participants' Mandarin proficiency and agreement with the exposure and usage

	7-point Likert scale		5-point Likert scale	
	Listening	Reading	Exposure	Usage
L1 TM (14)	*5.79 ± 1.05*	*5.86 ± 1.01*	*4.93 ± 0.28*	*4.93 ± 0.28*
L1 TM-Min (21)	*6 ± 1.02*	*5.76 ± 1.34*	*4.67 ± 0.97*	*4.67 ± 0.91*
L1 TM-Hakka (2)	*5 ± 1.41*	*5 ± 1.41*	*4.5 ± 0.71*	*4.5 ± 0.71*
Total (37)	*5.86 ± 1.13*	*5.76 ± 1.23*	*4.76 ± 0.76*	*4.76 ± 0.72*

Note. Within parentheses is the number of participants

All participants reported using Mandarin as their dominant language on a daily basis (Average = 87%). 21 (56.7%) out of 37 participants reported Taiwanese Min also as their native language (L1), a 7-point Likert scale was used to assess their proficiency (*Mean* = 4.05 ± 1.5). Additionally, 2 participants (5.4%) reported Hakka as their L1 with low proficiency (*Mean* = 1.5). Among 23 bilinguals, 22 of them reported using Taiwanese Min in their day-to-day life (Average = 14%) with 1 participant using Hakka 5%. Last, 9 participants (24.3%) reported using English or other languages daily (Average = 17.8%).

5. Discussion

The present research examined whether TM listeners could anticipate the *wh*-question and declarative according to BM prosodic cues in pre-*wh*-word and non-*wh*-counterpart contour. Our results demonstrated that TM listeners faced challenges anticipating the clause types in comparison to the findings of BM listeners in Gryllia et al's (2020) study, even though people from these two regions can usually communicate without difficulty.

Based on the above results, our research questions in section 1 have been addressed: First, TM listeners demonstrated the ability to anticipate the clause type based on BM prosody. The significant effect of the intended clause type was observed from Gate II onwards. Second, the declarative clause type was less challenging among all gates. Though there was no significant effect of the intended clause type in Gate I, participants performed slightly above the chance level in declaratives (53.6%) compared to *wh*-questions (46.4%). Third, TM listeners' responses were affected by the Verb tones, and the association between intended clause types and tone in Gate III was significant, in particular when the Verbs bore T1, T2, and T4, whereas T3 displayed different distributions.

According to Gryllia et al. (2020), the prosodic markings in Gate I and Gate II shared similar features in the *wh*-questions. First, a higher F0 maximum (H) was produced in S2 and S4 which both bore T1; second, the duration of S1 and S2 in Gate I, as well as S3 and S4 in Gate II were shorter compared to declaratives. However, these prosodic cues did not seem to facilitate anticipating *wh*-questions for TM listeners, the accuracy in both gates was below the chance level (Gate I: 46.4%, Gate II: 41.5%). Even in Gate III, the anticipation was only evident for T1 (62.3%), while for most other tones, it was not much better than the chance level.

During Gate I, participants struggled to anticipate any clause types when only a disyllabic Subject (e.g., proper name) was played. Moreover, 9 participants provided additional feedback about the difficulty of deciding on sentence continuations, as both options seemed equally valid for selection based on the information they heard. With the second disyllable word (i.e., Adverb) provided in Gate II, the anticipation in declarative clause type yielded the highest. In Gate III, we observed that participants tended to anticipate declaratives more often in T2, T3, and T4 than in T1, particularly in T3 and T4. Conversely, when the Verb tone bore T1 seemed to facilitate identifying *wh*-questions among other tones, which aligns with Gryllia et al. (2020); see **Figure 2** and **Table 4**. In addition, they mentioned that the maximum pitch of Verbs in T1 and T4 was higher, but only T1 improved anticipation of *wh*-questions.

From Gryllia et al.'s (2020) production study, they showed that there were prosodic differences between the two clause types. For instance, *wh*-questions exhibited a faster speech rate, higher F0, and shorter duration compared to declaratives. Though previous studies comparing BM and TM in sentence production found that TM had a slower tempo in declarative sentences (Tseng, 2004) and a lower F0 register (Torgerson, 2005; Chang, 2010), these prosodic features could not fully explain the enhanced accuracy in selecting declaratives for TM listeners. In fact, the declarative clause type seems to be more predictable than *wh*-questions in general, since BM listeners also anticipated declaratives relatively more often than *wh*-questions in Gryllia et al.'s (2020). However, it is evident that the anticipation of *wh*-questions remained highly unpredictable for TM listeners, even when more information was available in Gate III.

Therefore, conducting a supplementary production study on TM speakers using the same stimuli is essential to obtain prosodic properties of the two clause types in TM. This additional study will not only provide us with more insights into the similarities and differences compared to

BM but also enable a comprehensive interpretation of the current results. As evidenced by the association and effect of Verb tones on TM listeners' responses, the interaction between tone and intonation appears to influence anticipation, especially when the Verb bore T3. Chang (2010) found that T3 in isolation is realized as a low-dipping pattern in BM but as a low-falling pattern in TM. Moreover, the impact of T3 and T4 was reported in their perception study, where TM listeners faced difficulties identifying isolated T4 in BM and often perceived it as T3, and BM listeners faced the opposite misidentification. Although our task was not designed to examine the interaction by distinguishing the two clause types, our findings provided informative observations on tone effect in the pre-*wh*-word region and its counterpart.

Nevertheless, there were some limitations and further improvements that we came across during this study. Firstly, despite the advantages of online experiments in reaching a broader pool of participants, the less controllable environment compared to lab-setting could have influenced the results. For example, any disruption from the surrounding or malfunction of the device could have occurred, but we would have no information regarding it unless participants reported themselves. Secondly, while the marginally unbalanced responses for each gate and tone did not have a significant statistical impact, there is potential for further improvement in the current statistical analysis by using balanced data. Thirdly, the questionnaire could have been more precise in assessing participants' exposure to Mandarin, as individuals who were familiar with or frequently exposed to BM may have performed differently compared to those who did not. Lastly, though 56.7% of the participants reported Taiwanese Min as their L1, due to the low language usage (14%), we did not perform additional analysis of the effect, but future studies may need to take it into account as the cross-language effect may occur (e.g., Choi et al., 2016; Foltz, 2021).

Exploring how prosody shapes language processing and influences communication in diverse linguistic contexts continues to spark the interest of researchers in this ongoing and active field. Our results shed light on the prosodic anticipation of tone languages and enhanced understanding of the two Mandarin varieties. Moreover, new research inquiries have been raised. For example, whether TM speakers produce *wh*-questions with unique prosodic features compared to BM speakers, or whether the facilitation or interference of tones can be observed in reaction time. Furthermore, conducting experiments using methods focusing on temporal results gives us more detailed information about the real-time processing of the tone language.

6. Conclusion

Our findings supported the hypotheses and revealed that, despite lower overall performance compared to the participants from Beijing in Gryllia et al. (2020), Taiwanese listeners successfully anticipated the declarative clause type based on BM prosodic cues. Moreover, the tonal properties of the Verbs in Gate III influenced clause type anticipation and had a significant association. This study highlights the importance of conducting further production analysis to explore the prosodic markings of TM speakers.

References

Anwyl-Irvine, A.L., Massonié, J., Flitton, A., Kirkham, N.Z., & Evershed, J.K. (2019). Gorilla in our midst: an online behavioural experiment builder, *Behavior Research Methods.* https://doi.org/10.3758/s13428-019-01237-x

Balota, D. A., Pollatsek, A., & Rayner, K. (1985). The interaction of contextual constraints and parafoveal visual information in reading. *Cognitive Psychology, 17*(3), 364–390. https://doi.org/10.1016/0010-0285(85)90013-1

Bansal, S., Ford, J. M., & Spering, M. (2018). The function and failure of sensory predictions. *Annals of the New York Academy of Sciences, 1426*(1), 199–220. https://doi.org/10.1111/nyas.13686

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Chang, C. Y. (2010). *Dialect differences in the production and perception of Mandarin Chinese tones* (Doctoral dissertation). Retrieved from https://etd.ohiolink.edu/apexprod/rws_etd/send_file/send?accession=osu1282159730&disposition=inline

Chen, Y. (2022). Tone and Intonation. In C. Huang, Y. Lin, I. Chen, & Y. Hsu (Eds.), *The Cambridge Handbook of Chinese Linguistics* (Cambridge Handbooks in Language and Linguistics, pp. 336-360). Cambridge: Cambridge University Press. https://doi.org/10.1017/9781108329019.019

Choi, W., Tong, X., & Cain, K. (2016). Lexical prosody beyond first-language boundary: Chinese lexical tone sensitivity predicts English reading comprehension. *Journal of Experimental Child Psychology, 148*, 70–86. https://doi.org/10.1016/j.jecp.2016.04.002

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *The Behavioral and Brain Sciences*, *36*(3), 181–204. https://doi.org/10.1017/S0140525X12000477

Cutler, A., Dahan, D., & van Donselaar, W. (1997). Prosody in the comprehension of spoken language: A literature review. *Language and Speech, 40*(2), 141–201. https://doi.org/10.1177/002383099704000203

de Ruiter, J.P., Mitterer H., & Enfield, N. J. (2006). Projecting the end of a Speaker's Turn: A Cognitive Cornerstone of Conversation. *Language (Baltimore), 82*(3), 515–535. https://doi.org/10.1353/lan.2006.0130

Eberhard, D. M., Simons, G. F., & Fennig, C. D. (Eds.). (2023). *Ethnologue: Languages of the World*. Twenty-sixth edition. Dallas, Texas: SIL International. Retrieved from https://www.ethnologue.com/language/cmn/

Ehrlich SF, Rayner K. (1981). Contextual effects on word perception and eye movements during reading. *Journal of Verbal Learning and Verbal Behavior, 20*(6), 641–655. https://doi.org/10.1016/S0022-5371(81)90220-6

Ferreira, F., & Chantavarin, S. (2018). Integration and Prediction in Language Processing: A Synthesis of Old and New. *Current Directions in Psychological Science : a Journal of the American Psychological Society, 27*(6), 443–448. https://doi.org/10.1177/0963721418794491

Ferreira, F., & Qiu, Z. (2021). Predicting syntactic structure. *Brain Research*, 1770, 147632–147632. https://doi.org/10.1016/j.brainres.2021.147632

Garrod, S., & Pickering, M. J. (2015). The use of content and timing to predict turn transitions. *Frontiers in Psychology, 6*, 751–751. https://doi.org/10.3389/fpsyg.2015.00751

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. https://doi.org/10.5334/labphon.169

Hsu, Y.-Y., & Xu, A. (2020). Interaction of prosody and syntax-semantics in Mandarin wh-indeterminates. *The Journal of the Acoustical Society of America, 148*(2), EL119–EL124. https://doi.org/10.1121/10.0001676

Ito, K., & Speer, S. R. (2008). Anticipatory effects of intonation: Eye movements during instructed visual search. *Journal of Memory and Language, 58*(2), 541–573. https://doi.org/10.1016/j.jml.2007.06.013

Jackendoff, R. (2002). *Foundations of Language.* Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198270126.001.0001

Khlifat, J. A. (2016). The Interaction of Lexical Tone and Intonation in Mandarin. *International Journal on Studies in English Language and Literature (IJSELL). Volume 4*, pp 18-24. http://dx.doi.org/10.20431/2347-3134.0405003

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature (London), 307*(5947), 161–163. https://doi.org/10.1038/307161a0

Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience, 31*(1), 32–59. https://doi.org/10.1080/23273798.2015.1102299

Liu, M., Chen, Y., & Schiller, N. O. (2022). Context Matters for Tone and Intonation Processing in Mandarin. *Language and speech, 65*(1), 52–72. https://doi.org/10.1177/0023830920986174

Long, D. L., & Lea, R. B. (2005). Have We Been Searching for Meaning in All the Wrong Places? Defining the "Search After Meaning" Principle in Comprehension. *Discourse Processes, 39*(2-3), 279–298. https://doi.org/10.1080/0163853X.2005.9651684

Magyari, L., & de Ruiter, J. P. (2012). Prediction of turn-ends based on anticipation of upcoming words. *Frontiers in Psychology, 3*, 376–376. https://doi.org/10.3389/fpsyg.2012.00376

Mennen, I., & de Leeuw, E. (2014). Beyond Segments. *Studies in Second Language Acquisition, 36*(2), 183–194. https://doi.org/10.1017/S0272263114000138

Nakamura, C., Arai, M., & Mazuka, R. (2012). Immediate use of prosody and context in predicting a syntactic structure. *Cognition, 125*(2), 317–323. https://doi.org/10.1016/j.cognition.2012.07.016

Roll, M., Horne, M., & Lindgren, M. (2010). Word accents and morphology—ERPs of Swedish word processing. *Brain Research, 1330*(12 May), 114–123. https://doi.org/10.1016/j.brainres.2010.03.020

Stromswold, K., Eisenband, J., Norland, E., & Ratzan, J. (2002). Tracking the acquisition and processing of English passives: Using acoustic cues to disambiguate actives and passives. In *The 16th Annual CUNY Conference on Human Sentence Processing.* Cambridge, MA. Retrieved from http://ruccs.rutgers.edu/images/personal-karin-stromswold/publications/cuny.pdf & http://ruccs.rutgers.edu/images/personal-karin-stromswold/publications/Kstrompsv.pdf

Torgerson, R. C. (2005) *A Comparison of Beijing and Taiwan Mandarin Tone Register: An Acoustic Analysis of Three Native Speech Styles* (Master's thesis). Retrieved from https://scholarsarchive.byu.edu/etd/642

Tseng, C. 2004. Prosodic Properties of Intonation in Two Major Varieties of Mandarin Chinese: MainlandChina vs. Taiwan. *Proceedings of the Intonational Symposium on Tonal Aspect of Languages: with Emphasis onTone Languages,* p.189-191. Beijing, China. Retrieved from https://sprosig.org/tal2004/tal2004-Beijing/Tseng.pdf

Van Petten, C., & Luka, B. J. (2012). Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology, 83*(2), 176–190. https://doi.org/10.1016/j.ijpsycho.2011.09.015

Wang, T., Liu, J., Lee, Y., & Lee, Y. (2020). The interaction between tone and prosodic focus in Mandarin Chinese. *Language and Linguistics (Taipei), 21*(2), 331–350. https://doi.org/10.1075/lali.00063.wan

Xu, Y. 2004. Understanding tone from the perspective of production and perception. *Language and Linguistics 5*(4). 757–797. Retrieved from http://www.homepages.ucl.ac.uk/~uclyyix/yispapers/XUYI_Lang_Ling.pdf

Xu, Y. (2019). Prosody, tone, and intonation. In *The Routledge Handbook of Phonetics* (1st ed., pp. 314–356). Routledge. https://doi.org/10.4324/9780429056253-13

Yang, C. (2016). *The acquisition of L2 Mandarin prosody.* Vol. 1. Netherlands:John Benjamins Publishing Company. https://doi.org/10.1075/bpa.1

Yang, Y. (2018). *The two sides of Wh-indeterminates in Mandarin: a prosodic and processing account* (Doctoral dissertation). Retrieved from https://scholarlypublications.universiteitleiden.nl/access/item%3A2952546/view

Yuan, J. (2006). Mechanisms of Question Intonation in Mandarin. *Chinese Spoken Language Processing*, 19–30. Berlin, Heidelberg: Springer Berlin Heidelberg. 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. https://doi.org/10.1121/1.3651818

Zeng, X., Martin, P., & Boulakia, G. (2004). Tones and Intonation in Declarative and Interrogative Sentences in Mandarin. Retrieved from https://sprosig.org/tal2004/tal2004-Beijing/Zeng-Martin.pdf

Zou, T., Chen, Y. &, Caspers, J. (2015). Attention redistribution and segment-tone integration in Mandarin tone acquisition by L2 learners. *Proceedings of the 18th International Congress of Phonetic Sciences*. https://hdl.handle.net/1887/35499

Appendix A

1. Response Stimuli (in traditional and simplified character scripts, examples from p.12):

	Wh-question continuation				
	zuótiān	*tōule*	***shénme***	*gěi*	*báibīng*
Traditional	昨天	偷了	**什麼**	給	白冰
Simplified	昨天	偷了	**什么**	给	白冰
	yesterday	steal.PERF	**what**	for	Baibing
	'stole what for Baibing yesterday?'				

	Declarative continuation				
	zuótiān	*tōule*	***nízi***	*gěi*	*báibīng*
Traditional	昨天	偷了	**呢子**	給	白冰
Simplified	昨天	偷了	**呢子**	给	白冰
	yesterday	steal.PERF	**woolen**	for	Baibing
	'stole woolen for Baibing yesterday.'				

2. Instruction

2.1. Practice phase

此實驗是聽音檔預測接下來的句子類型，以漸進式的方式進行。第一階段您只會聽到一個詞；第二階段兩個詞...等。請根據聽到的內容，選擇一個您認為合適的句子，選項有兩個：陳述句和疑問句。

'This experiment is listening to a recording and predicitng the type of sentence that will follow, it proceeds progressively. In the first stage, you will only hear a word; the second stage is two words…etc. Please based on the information you listen to, choose a sentence that you think is appropriate. there are two options: declarative and question sentence.'

接下來是一個簡單的練習，請點選 "OK" 開始。

'Next is a simple exercise, please press "OK" to start.'

1.2. Test phase

1.2.1 Gate I

- First slide:

第一階段的聽力測驗即將開始。

‘The first stage of the listening test is about to start.’

螢幕中顯示的句子用詞為中國用詞，比如：提子=葡萄、殼子=貝殼、呢子=羊毛、皮子=皮革；郵=郵寄、淘=洗…等。這些用詞差異並不影響您的作答。

‘The words in the sentence shown on the screen are words that used in China, for example, *tízi =pútáo ‘grape’, kézi =bèiké ‘shell’, nízi =yángmáo ‘woolenl’, pízi =pígé ‘leather’; yóu =yóujì ‘to mail’, táo =xǐ ‘to wash’*…etc. These differences do not affect your answer.’

請仔細閱讀以下說明。

‘Please read the following instruction carefully.’

- Second slide:

這是一個聽音檔完成句子的實驗。音檔中是一個句子的片斷，只有<u>一個詞</u>(人名)。因為句子不完整，所以請根據聽到的訊息，選擇該句子應該如何繼續說才完整。選項有兩個，一個陳述句，一個疑問句，請通過滑鼠點擊您認為正確的選項。完成一題後，請點擊"OK"繼續。

‘This is an experiment in which you listen to a recording to complete a sentence. The recording is a sentence fragment with only <u>one word</u> (person’s name). Since the sentence is incomplete, based on the information you hear, choose how the sentence should be said further to be complete. There are two options, one declarative and one interrogative sentence, click on the option you think is correct. When you have completed a question, click “OK” to continue.’

1.2.2. Gate II

這是一個聽音檔完成句子的實驗。音檔中是一個句子的片斷，只有兩個詞。因為句子不完整，所以請根據聽到的訊息，選擇該句子應該如何繼續說才完整。選項有兩個，一個陳述句，一個疑問句，請通過滑鼠點擊您認為正確的選項。完成一題後，請點擊“OK”繼續。

‘This is an experiment in which you listen to a recording to complete a sentence. The recording is a sentence fragment with only two words. Since the sentence is incomplete, based on the information you hear, choose how the sentence should be said further to be complete. There are two options, one declarative and one interrogative sentence, click on the option you think is correct. When you have completed a question, click “OK” to continue.’

1.2.3. Gate III

這是一個聽音檔完成句子的實驗。音檔中是一個句子的片斷。因為句子不完整，所以請根據聽到的訊息，選擇該句子應該如何繼續說才完整。選項有兩個，一個陳述句，一個疑問句，請通過滑鼠點擊您認為正確的選項。完成一題後，請點擊"OK"繼續。

‘This is an experiment in which you listen to a recording to complete a sentence. The recording is a sentence fragment. Since the sentence is incomplete, based on the information you hear, choose how the sentence should be said further to be complete.

There are two options, one declarative and one interrogative sentence, click on the option you think is correct. When you have completed a question, click “OK” to continue.’

3. Questionnaire

1. 請問您的生理性別：
   ‘Please indicate your birth sex:’
   - 男性 ‘Male’
   - 女性 ‘Female’
   - 不提供 ‘Not provided’
2. 請問您的年齡：
   ‘Please indicate your age:’
3. 您有無任何聽力障礙：
   ‘Whether you have any hearing impairment:’
   - 有 ‘Yes’
   - 無 ‘No’
4. 您有無任何閱讀障礙或其他語言障礙的醫療診斷：
   ‘Whether you have any medical diagnosis of dyslexia or other language disorders:’
   - 有 ‘Yes’
   - 無 ‘No’
5. 您的中文閱讀能力：
   ‘Your Chinese reading skills:’
   - 很差 ‘very poor’
   - 勉強 ‘poor’
   - 尚可 ‘acceptable’
   - 普通 ‘average’
   - 還不錯 ‘not bad’
   - 很好 ‘good’
   - 非常好 ‘excellent’
6. 您的中文口語理解能力：
   ‘Your Chinese listening comprehension skills:

- 很差 ‘very poor’
- 勉強 ‘poor’
- 尚可 ‘acceptable’
- 普通 ‘average’
- 還不錯 ‘not bad’
- 很好 ‘good’
- 非常好 ‘excellent’

請選擇適用您的選項：

‘Please select the option that applies to you:’

7. 中文是我居住環境的主要語言

   ‘Chinese is the primary language of my living envrionment’
   - 非常不同意 ‘Strongly disagree’
   - 不同意 ‘disagree’
   - 普通 ‘neither agree or disagree’
   - 同意 ‘agree’
   - 非常同意 ‘strongly agree’
8. 中文是我日常接觸最多的語言

   ‘Chinese is the languae I have the most contact with on a daily basis’
   - 非常不同意 ‘Strongly disagree’
   - 不同意 ‘disagree’
   - 普通 ‘neither agree or disagree’
   - 同意 ‘agree’
   - 非常同意 ‘strongly agree’
9. 除了中文，我還使用其他母語：

   ‘In addition to Chinese, I also use other native language:’
   - 無 ‘None’
   - 台語 ‘Taiwanese Min’
   - 客家語 ‘Hakka’
   - 其他，請說明：‘Other, please indicate: ’

10. 以日常生活來看，請具體說明您語言使用的百分比？（合在一起 100%，比如：中文 60%，台語 40%）

‘In terms of daily life, please specify the percentage of languages you use? (Combined to 100%, e.g., Chinese 70%, Taiwanese Min 40%)’

Appendix B

Table 1. Best-fitting models

	AIC	BIC	logLik	deviance	Chisq	Df	Pr(>Chisq)
Gate I.m0	1909.3	1925.2	-951.66	1903.3	N/A	N/A	N/A
Gate II.m1	1789.6	1810.8	-890.81	1781.6	37.15	1	< .001
Gate III.m2	1830.7	1867.8	-908.37	1816.7	67.91	3	< .001
Tone.m1	1873.7	1905.5	-930.87	1816.7	65.57	3	< .001
allGate.m3	5611.8	5662.9	-2797.9	5595.8	27.25	2	< .001
Gate.m1	5863.8	5715.8	-2836.9	5673.8	48.004	2	< .001