Listeners do not use Mandarin Chinese tone sandhi to predict upcoming words: a visual world eye-tracking study

Yiling Huo[*] Wing Yee Chow[*]

September 13, 2025

Abstract

Listeners use rich information to predict upcoming language on the fly, but it is less clear to what extent they can use phonological information to predict upcoming words. In this study, we investigate the role of phonological information as an input to lexical predictions by capitalising on two right-dominant tone sandhi patterns in Mandarin Chinese, the T3 sandhi and the *yi* sandhi. In both cases, the tone of the first syllable in the sandhi domain changes systematically depending on the tone of the syllable that follows. In two visual world eye tracking experiments (Experiments 1 and 3), participants viewed pairs of objects while listening to non-constraining sentences that identified one of the objects on the visual display. The head noun was always preceded by a numeral and a classifier, and the numeral may undergo tone sandhi depending on the tone of the classifier (e.g., *yi4 zhang1 chuang2*, one $\mathrm{CL}_{\mathrm{zhang}}$ bed). We manipulated the tone of the numeral to be either informative about the target's identity (Different Tones condition), or uninformative (Same Tones condition). While we report suggestive evidence for listeners using T3 sandhi cues, but not *yi* sandhi cues, to generate lexical predictions Experiment 1, neither type of tone sandhi cues was found to impact listeners' predictions in our subsequent replication study (Experiment 3). An analysis of data from both experiments using Bayesian principles suggested that listeners did not use informative tone sandhi cues to compute lexical predictions. Meanwhile, results from two acceptability judgment tasks (Experiments 2 and 3) showed that listeners have robust linguistic knowledge of both tone sandhi patterns. Overall, we found that listeners, despite being sensitive to tone sandhi, could not use them to make lexical predictions on the fly. Set against previous evidence for the listeners' impressive ability to predict using a wide array of cues, the present results suggest a limit to listeners' capacity to use phonological cues for prediction.

Keywords sentence processing; prediction; lexical prediction; tone sandhi; eye-tracking; visual world paradigm

1 Introduction

Psycholinguists have provided compelling evidence that comprehenders make real-time predictions about upcoming language input during comprehension (Kuperberg & Jaeger, 2016; Pickering & Gambi, 2018; Ryskin & Nieuwland, 2023; Van Petten & Luka, 2012). While a wide range of linguistic and non-linguistic information has been shown to contribute to predictions during comprehension, it is less clear to what extent phonological information feed into predictions, especially lexical predictions, during comprehension. In this study, we investigate the role of phonological information in lexical prediction by examining tone sandhi patterns in Mandarin Chinese. Tone sandhi involves predictable changes in a syllable's lexical tone based on its phonological context, such that the presence of a syllable that has undergone tone sandhi can be informative about upcoming language input. As such, tone sandhi offers a good testing ground for understanding how listeners might use phonological information to generate predictions during sentence comprehension.

1.1 Prediction during sentence comprehension

The computation of prediction has been proposed as a universal principle underlying neural processes (Bar, 2011). In language comprehension, it is well-established that listeners predict upcoming linguistic input in real time. A classic visual world eye-tracking study by Altmann & Kamide (1999) demonstrated this by showing that when listeners heard sentences like "The boy will eat/move the cake," they were more likely to look anticipatorily

[*] Division of Psychology and Language Sciences, University College London, London, UK

at edible objects (e.g., cake) after hearing the verb “eat” compared to a less restrictive verb such as “move.” This indicates that listeners could use the verb’s selectional restrictions to anticipate an upcoming noun.

Over the past three decades, a large body of research has shown that comprehenders’ predictions about upcoming words can be guided by various types of information, including both linguistic and real-world knowledge. For example, studies have demonstrated that linguistic markers such as tense (Altmann & Kamide, 2007), gender (Lew-Williams & Fernald, 2007, 2010; Stone, Verı́ssimo, et al., 2021), case (Kamide, Scheepers, et al., 2003), number (Lukyanenko & Fisher, 2016), and noun class (Chow & Chen, 2020; Kwon et al., 2017) can feed predictions. Additionally, sentential and discourse context (Altmann & Kamide, 1999; Federmeier & Kutas, 1999; Kutas & Hillyard, 1984; Otten & Van Berkum, 2008; Van Berkum et al., 2005) and event knowledge (Chow et al., 2016; Kamide, Altmann, et al., 2003) have also been shown to play important roles in the computation of predictions during comprehension.

However, relatively little is known about whether and how phonological information may feed into predictions during real-time language comprehension. In the present study, therefore, we aim to investigate whether listeners use informative phonological cues to compute lexical predictions.

1.2 Phonological information as an input to language prediction

1.2.1 Lexically- and morphologically-informative phonological information

Can listeners use phonological information they encounter before the target to generate predictions about upcoming linguistic content? This question has been investigated at different levels of linguistic representations. One of the reasons why different levels of linguistic representations are involved is that some phonological cues are informative about upcoming language input directly on the lexical or morphemic level.

For instance, many languages use contrastive pitch accent to express information structure in the sentence or discourse. In English, a sentence such as “*KATIE did not win a truck*”, where the subject carries a contrastive accent L + H*, is highly likely to be continued with another person who won a truck (“*..., LAURA did*”); whereas a sentence such as “*Katie did not win a TRUCK*”, where the accent was on the object, is likely continued with an alternative item that Katie won (“*..., she won a MOTORCYCLE*”) (K. Ito & Speer, 2008). K. Ito & Speer (2008) showed that listeners can use emphatic or contrastive pitch accents in English adjectives to predict upcoming nouns. They presented participants with a display of ornaments varying in shape and colour, while the participants listened to instructions to pick out the ornaments. Eye movements suggested that participants moved their eye gaze to the target object faster when the two-part instruction carried a contrastive pitch accent that was congruent with the contrast (“*Please find the blue ball. Now, please find the GREEN ball*”), compared to when the pitch accent was neutral (“*... the green ball*”). These results suggest that listeners used the congruent contrastive pitch accent to anticipate the identity of upcoming nouns.

Weber et al. (2006) also examined how contrastive pitch accents influence comprehenders’ visual attention. Participants viewed displays that were either contrastive (e.g., an array containing two pairs of scissors in different colours) or non-contrastive displays (e.g., an array of four distinct objects) while listening to instructions with contrastive pitch accents (e.g., “*Please hand me the RED scissors/vase*”). Eye-tracking data showed that listeners made anticipatory fixations on the target only in contrastive displays, not in the non-contrastive ones. These findings suggest that listeners use contrastive pitch accents to quickly infer the speaker’s intent to specify a target within a contrastive pair.

Cross-linguistically, phonological cues can also be informative about higher-level language input a more local scale (e.g. on the morphemic level). For example, Roll et al. (2010) and Söderström et al. (2012) showed that listeners of Central Swedish use word accents to anticipate upcoming morphemes. In Central Swedish, the tone carried by a word’s stem is determined by which suffix is attached to it. For instance, noun stems such as *hatt* carries a low tone in *hatt-en* “hat_{SG}”, and a high tone in *hatt-ar* “hat_{PL}”. Similarly, verbs such as *läk* “heal” carries a low tone in the present tense form *läk-er*, but a high tone in the past tense form *läk-te*. Therefore, the tone on the stem provides direct cues to which suffix the word will have. In an ERP study, Roll et al. (2010) found that suffixes that mismatched with a word stem’s tone elicited greater brain responses compared to suffixes that matched with the tone. Similarly, Söderström et al. (2012) used a behavioural task in which participants determined a verb’s tense, and observed that stem tones that mismatched with the tense suffix led to slower response times than those that matched. Together, these studies provide converging evidence that word accents in Central Swedish can be used by listeners to anticipate an upcoming morpheme.

1.2.2 Phonologically-informative phonological information

While phonological cues such as contrastive pitch accent or Central Swedish word accent are informative about upcoming language input on a higher level than phonology, such as indicating the repetition of a previously named noun or indicating a specific morpheme, there are many phonological cues that are only informative on the phonological level. For instance, in English, the coronal nasal /n/ undergoes assimilation when preceding a bilabial (e.g. *green* [grim] *beans*, where the nasal sound /n/ becomes [m] before /b/). As such, an assimilated nasal sound can indicate an upcoming bilabial.

In the case of such phonological cues, one of their roles in language processing is to allow listeners to anticipate a particular phoneme. For example, Gow Jr (2001) presented participants with phrases that could undergo phonological assimilation (e.g. *green* [grim] *beans*). The coronal nasal /n/ in the first word either underwent appropriate assimilation ([n] - [m]), or no assimilation ([n] - [n]). Listeners were asked to monitor for bilabial sounds (e.g. [b] in *beans*) and press a button as soon as they heard one of them. Results showed that appropriate assimilation significantly reduced the time needed to recognise the trigger phonemes compared to no assimilation. This indicates that the recognition of the trigger phonemes were facilitated by an appropriate nasal assimilation in the preceding sound. One possible source of this facilitation is that upon hearing a nasal assimilation, listeners could quickly use this as a predictive cue to anticipate an upcoming bilabial consonant that triggered the assimilation (Gow Jr, 2001).

The anticipation of an upcoming phoneme can then facilitate word recognition. In a visual world eye-tracking study, Schreiber & McMurray (2019) showed that listeners could use coarticulatory information in the fricatives /s/ and / / to anticipate whether the upcoming vowel was rounded (/u/) or unrounded (/i/). Crucially, such anticipation resulted in anticipatory eye movements to the target object (e.g. *soup* or *seed*) when the fricative was correctly coarticulated compared to when it was incorrectly coarticulated. This shows that the anticipation of the vowel led to quicker identification of the word.

Interestingly, similar facilitatory effects have also been observed *across word boundary.* Salverda et al. (2014) showed that listeners could use coarticulatory information in a previous word to facilitate the recognition of the *upcoming word.* Similarly, Gow & McMurray (2007) presented listeners with visual displays containing four objects, each corresponding to possible two-word combinations in which the first word ended with either a coronal or a non-coronal consonant and the second word began with either a bilabial or a non-bilabial consonant (*green/swamp boat/dog*). Participants heard spoken instructions such as "*select the green boat*" and were asked to click on the appropriate object. Crucially, the coronal phoneme followed by the bilabial phoneme was either correctly assimilated (e.g. *select the green* [grim] *boat*) or not assimilated (*select the green* [grin] *boat*). Eye movement results suggested that listeners fixated on the bilabial-initial target object (*boat*) earlier when it was preceded by a correctly assimilated coronal ([grim]) than by a non-assimilated coronal ([grin]). These findings indicate that listeners anticipated the bilabial phoneme upon hearing the assimilated coronal and were able to identify the target lexical item (*boat*) more efficiently.

Cross linguistically, studies have also found similar cross-word-boundary facilitation effects of informative phonological cues. In the Kansai dialect of Japanese, a word can either be accented or unaccented. Accented words always starts with a low tone and rises on the accented mora, before dropping again (*hiKOoki* aeroplane, where capital letters indicate high-pitched moras) while an unaccented word can either have an all-high tone (*HITUJI* sheep) or an all-low tone (*simasima-no* striped). Crucially, when an all-low-tone modifier precedes a low-initial accented word, the final mora of the modifier is realised in a high tone instead of a low tone (*simasima-**NO** hiKOoki*). As a result, the pitch of the final mora can act as a cue to the pitch accent of the upcoming word. While in standard Tokyo Japanese, such predictive tonal changes do not exist, as unaccented words never starts with a high tone.

In their study, A. Ito & Hirose (2024) presented Kansai Japanese speakers with visual displays of objects featuring different colour patterns (e.g., striped or dotted sheep and airplanes), alongside a spoken phrase identifying one of the objects (e.g., *simasima-no HITUJI* "striped sheep"). The phrases were presented such that a pause of 800ms was inserted between the modifier and the target noun. Crucially, the two types of objects' labels (*HITUJI* "sheep" vs. *hiKOoki* "aeroplane") carry different initial pitch. In an informative condition, the modifier preceding the target object (*simasima-no* "striped") was low-pitch, meaning that its final mora changes pitch depending on which object would be the target noun (*simasima-**NO** hiKOoki* "striped airplane" vs. *simasima-**no** HITUJI* "striped sheep"), rendering the final mora of the modifier informative about the target noun's identity. In contrast, in an uninformative condition, the modifier was all-high-pitch (*MIZUTAMA-NO* "polka-dotted"), which means that the pitch of the modifier never changes regardless of whether a high-initial or a low-initial noun would follow (*MIZUTAMA-NO hiKOoki* "polka-dotted aeroplane" vs. *MIZUTAMA-NO HITUJI* "polka-dotted sheep"), rendering the modifier uninformative about the target noun's identity. Participants were instructed to select the target picture as quickly as possible. Reaction time data showed that listeners were faster to select the target when the modifier ended with an informative low tone (e.g., *simasima-no* (*HITUJI*)), indicating that informative low tones can facilitate the recognition of the upcoming noun.

These cross-word-boundary facilitation effects seem to suggest a possibility that comprehenders can not only anticipate upcoming phonological input, but also predict an upcoming *word* based on these phonological cues, at least in the visual world eye-tracking paradigm where there is a finite set of possible targets. The phonological cues lead to the prediction of an upcoming phonological form (a phoneme such as /b/ or a high tone), which subsequently becomes a prediction of a lexical item (*boat* or *HITUJI*).

However, studies such as Gow & McMurray (2007) and A. Ito & Hirose (2024) cannot provide substantial evidence that supports this possibility. Firstly, both studies found anticipatory eye movements or responses *after* the target word onset, not before. Thus it remains unclear whether the observed facilitation effects stem from pre-activation of the target words. Secondly, some effects were not found for all types of informative cues: in A. Ito & Hirose (2024)'s study, a facilitation effect was found for informative low tones (*simasima-**no***), but not for high tones that were equally informative (*simasima-**NO***). Finally, A. Ito & Hirose (2024) found the same facilitation effect of informative low tones in a control group of Standard Japanese speakers who had no previous exposure to the Kansai dialect. The same facilitation effects in the non-Kansai speakers suggest that the effects found in A. Ito & Hirose (2024) may stem from mechanisms that are independent of prediction. Taken together, these findings suggest that current evidence does not conclusively demonstrate that listeners generate lexical predictions based on phonological cues that are informative only at the phonological level.

In this study, therefore, we aim to further investigate whether comprehenders use informative phonological cues to predict upcoming words, capitalising on Mandarin Chinese tone sandhi—change in lexical tones induced by neighbouring tonal environments.

It is important to note that the current study focuses on phonological cues as an *input* to lexical prediction, and should be distinguished from the body of research that investigates whether phonological form is part of the *output* of predictions (DeLong et al., 2005; A. Ito et al., 2018; A. Ito, 2024; Kukona, 2020; Li et al., 2022; Martin et al., 2013; cf. Nieuwland et al., 2018) (for a review, see A. Ito, 2024). Although this line of research addresses an important question about what are being predicted during language comprehension, its findings do not speak to whether informative phonological cues can shape comprehenders' lexical predictions.

1.3 Using Mandarin Chinese tone sandhi to study phonological information in prediction during sentence comprehension

1.3.1 Mandarin Chinese tone sandhi patterns

Mandarin Chinese is a tonal language with four citation tones: T1 (high-level), T2 (high-rising), T3 (low/dipping), and T4 (falling). These lexical tones are syllable-level pitch contours that encode meanings (Chao, 1968). For example, *ma* in T1 (妈) means 'mother', in T2 (麻) it means 'hemp', *ma3* (马) 'horse', *ma4* (骂) 'to scold'. In Mandarin Chinese, as in many tonal languages, a syllable can undergo tone alternation without affecting its meaning. These lexical tone alternations are conditioned by adjacent tones, prosodic environment, or morphosyntactic context, a phenomenon known as tone sandhi (Zhang, 2014).

In Mandarin Chinese, where there are two successive T3 syllables, the first T3 syllable is not realised as a low/dipping tone, but as a high-rising tone that is perceptually indistinguishable from a T2 (Peng, 2000). This pattern is known as the T3 sandhi (W. S. Wang & Li, 1967). For example, when the adjective *xiao3* "small" modifies the noun *mao1* "cat", the phrase is pronounced *xiao3 mao1* "small cat/kitten" (*xiao3* keeps its low-dipping T3), yet when the adjective modifies a noun such as *gou3* "dog", the phrase is realised as *xiao2 gou3* "small dog/puppy" (*xiao3* changes to the high-rising T2[1]). Importantly, the T3 sandhi is a right-dominant phonological pattern; that is, the right side of the sandhi domain (the second T3 syllable) is preserved, while the left side (the first T3 syllable) is altered. This property means that the tone of one syllable can potentially be informative about the tone of the syllable that follows. In other words, hearing a typically T3 syllable in a high-rising T2 (e.g. 小 *xiao3* → *xiao2*) indicates that the T3 sandhi has taken place, and consequently that the next syllable is likely in T3 (e.g. *gou3*). Likewise, hearing *xiao3* in its base tone can be taken to suggest that the next syllable is not in T3.

Another tone sandhi pattern in Mandarin Chinese involves the morpheme/word *yi1* ("a, one"). When used as the indefinite article or a cardinal numeral, *yi1* is realised in the falling tone (T4) when followed by a syllable in T1, T2, or T3 (*yi4ban1* "normally, generally"; *yi4tong2* "together, at the same time"), but it is realised in

[1] In phonetics, the result of a T3 sandhi is sometimes denoted as a sandhi T3 or a T2-star (T2*). In this paper, for simplicity, we denote a sandhi T3 simply as T2, as there is evidence that shows that sandhi T3/T2* and original T2 are virtually indistinguishable in perception (Peng, 2000), and the distinction between original T2 and sandhi T3/T2* is not part of this study's research question.

the rising tone (T2) when followed by a T4 syllable (*yi2ding4* "certainly, must") (C. B. Wang, 2014). This is known as the *yi* sandhi.

Similar to the T3 sandhi, the *yi* sandhi is also right-dominant, since the left side of the sandhi domain (*yi*) changes its tone according to the syllable on the right (that is, the syllable following *yi*). This means that the *yi* sandhi can also serve as a predictive cue to the upcoming syllable's tone. However, different from the T3 sandhi, the *yi* sandhi is morphologically conditioned, in that it only applies to the word/morpheme *yi*, while the T3 sandhi applies to all syllables in T3.

In this study, we exploit the T3 sandhi as well as the *yi* sandhi in Mandarin Chinese numerals to test whether listeners use tonal cues to generate predictions about upcoming words.

1.3.2 Tone sandhi as a predictive cue in Mandarin Chinese noun phrases

In Mandarin Chinese, when a noun is modified by a numeral or a demonstrative, a classifier must occur after the modifier and before the noun (*yi4 ben3 shu1*, one $\mathrm{CL}_{\mathrm{ben}}$ book, "a book/one book"; *yi2 shan4 men2*, one $\mathrm{CL}_{\mathrm{shan}}$ door, "a door/one door") (C. N. Li & Thompson, 1989). The majority of classifiers in Mandarin Chinese are only used with a specific class of nouns. For example, the classifier *tiao2* can be only used with things that are long and thin, such as *wei2jin1* (scarf) and *lu4* (road); while the classifier *zhi1* can only be used with animals that are relatively small in size such as *tu4zi* (rabbit) or *qing1wa1* (frog), but not *niu2* (cow) or *da4xiang4* (elephant), which require another classifier *tou2*. As such, these specific classifiers in Mandarin Chinese pose constraints on what class of nouns can follow.

Previous research has shown that classifiers and nouns are closely associated during sentence comprehension, such that comprehenders can quickly use a classifier to predict an upcoming noun (J. Chen et al., 2010; Grüter et al., 2018; Huettig et al., 2010; Klein et al., 2012); and when a noun is activated during language comprehension, the classifier associated with it can also be activated (Chow & Chen, 2020; Kwon et al., 2017).

Before the classifier, crucially, both the T3 sandhi and the *yi* sandhi can apply to the numeral. The *yi* sandhi affects the numeral/indefinite article *yi1* (一, "a, one"), while T3 sandhi affects numerals in T3, such as *liang3* (两, "two") and *wu3* (五, "five"). In Mandarin Chinese noun phrases, a classifier must follow the numeral[2], and classifiers are closely associated with specific noun classes. As a result, the tone of numerals like *yi1* or *liang*—as altered by sandhi rules—can provide cues about the tone of the upcoming classifier and, by extension, the identity of the following noun. For example, hearing *yi4* (with a falling tone) suggests that the classifier begins with tones 1, 2, or 3, while *yi2* (with a rising tone) signals a T4 classifier—and therefore a noun that is compatible with such classifiers. Similarly, *liang3* indicates a classifier in T1, T2, or T4, whereas *liang2* points to a classifier in T3. In this way, the surface tone of a numeral can serve as an early phonological cue to the structure and semantics of the noun phrase.

A recent study by Liu et al. (2025) proposed that Mandarin Chinese speakers can use tone sandhi information to predict upcoming words. They employed the visual world eye-tracking paradigm to ask whether listeners can use the *yi* sandhi in the indefinite article *yi* to predict an upcoming classifier and noun. In each trial, participants listened to a question such as 请问有一张邮票在左边吗? "Is there one $\mathrm{CL}_{\mathrm{zhang}}$ stamp on the left (of the screen)?" while viewing a visual display of two objects (the target and a competitor). The auditory stimuli were presented at a fixed rate of 800ms per syllable (SOA = 800ms) and the critical noun phrase always consisted of the numeral *yi*, a classifier, and the target noun. Eye movement data showed that listeners were faster to look at the target object when the tone of *yi* was informative about the classifier's tone than when it was uninformative. These findings provided suggestive evidence for the use of tone sandhi in lexical prediction, but it remains unclear whether they will generalise to other settings, for example when the stimuli are presented in a more natural and realistic speech rate. Moreover, the fact that this study always used the same carrier sentence with *yi* as the numeral and had no filler sentences may have enhanced the salience of the informativeness of *yi*, potentially allowing participants to develop experiment-specific strategy to pick out the target based on the tone of *yi*. As such, more evidence is needed to establish whether Mandarin Chinese listeners can use tone sandhi cues to predict upcoming lexical items during naturalistic spoken sentence comprehension.

[2] Mandarin Chinese does not normally allow adjectives between the numeral and the classifier, with the only exceptions being *da4* "big" and *xiao3* "small" for a limited set of nouns. For example, *yi2* ***da4*** *zhi1 zhang1lang2* (one big $\mathrm{CL}_{\mathrm{zhi}}$ cockroach, "a big cockroach") where the adjective "big" is put in-between the numeral and the classifier is possible for some speakers; although *yi4 zhi1* ***da4*** *zhang1lang2* (one $\mathrm{CL}_{\mathrm{zhi}}$ big cockroach, "a big cockroach") is more canonical and more frequent, where the numeral is followed by the classifier directly and the adjective is inserted after the classifier.

1.4 The present study

In the present study, we use Mandarin Chinese tone sandhi to investigate whether phonological information can serve as an input to lexical prediction during language comprehension. Across three experiments, we aim to determine whether listeners use tone sandhi cues in a numeral to predict the upcoming classifier and noun during spoken sentence comprehension and whether there is a correlation between a listener's sensitivity to tone sandhi and their ability to use this information for prediction. Specifically, we tested two tone sandhi patterns in Mandarin Chinese: the T3 sandhi and the *yi* sandhi.

To preview, Experiment 1 employed the visual world eye-tracking paradigm to explore whether listeners use tone sandhi information in real-time lexical prediction. We used a naturalistic speech rate, varied the sentence context in the stimuli, and included filler items to prevent participants from developing experiment-specific strategies. Experiment 2 followed up on the findings from Experiment 1 by using an acceptability rating task to assess listeners' sensitivity to violations of both T3 sandhi and *yi* sandhi. By examining listeners' acceptability ratings of stimuli that either complied with or violated these tone sandhi rules, this experiment sought to provide insights into comprehenders' linguistic knowledge about tone sandhi. Experiment 3 was a direct replication of both Experiment 1 (eye-tracking) and Experiment 2 (acceptability rating) within the same group of participants. This allowed us to correlate individual participants' sensitivity to tone sandhi with their ability to use it for prediction during real-time sentence comprehension, providing a clearer picture of the relationship between listeners' sensitivity and their use of tone sandhi for prediction.

2 Experiment 1

In the present experiment, we adopted the visual world eye-tracking paradigm to investigate whether native Mandarin Chinese speakers use their knowledge of tone sandhi (specifically, the *yi* sandhi and the T3 sandhi) to predict upcoming words. We tracked participants' eye movements while they listened to sentences where the tone of a numeral was either informative or uninformative about the identity of an upcoming classifier and noun.

Participants were presented with a pair of objects (e.g. *yi4 zhang1 chuang2* one CL_{zhang} bed "one bed", *yi2 shan4 men2* one CL_{shan} door "one door"), and their eye movements were recorded as they listened to a sentence that identified one of the objects with a critical noun-phrase (NP) consisting of a numeral, a classifier, and the target noun (e.g. "Uncle Li saw one CL bed."). The spoken sentence always had a neutral context that was compatible with either object in the visual display. The objects were paired based on the nouns that they depict and the classifiers the nouns take. In the Different Tones (experimental) condition, one of the classifiers in the pair triggered a tone sandhi in the numeral that precedes it, thus, the numeral took on different tones depending on which object is to be named (*yi4 zhang1 chuang2* "one bed" vs. ***yi2*** *shan4 men2* "one door"). Therefore, the tone of the numeral was informative about the identity of the upcoming classifier and noun. In the Same Tones (control) condition, neither of the classifiers triggered a tone sandhi, thus the numeral always had the same tone regardless of which object was to be named (*yi4 zhang1 chuang2* "one bed" vs. *yi4 ba3 deng4zi* "one stool"). As a result, the numeral's tone was uninformative about the target's identity in the Same Tones condition.

Note that the objects in the visual displays were chosen such that the two nouns always took distinct classifiers. This means that the classifier was equally informative about the noun's identity across both conditions.

If listeners can use the tone of the numeral to anticipate the rest of the noun phrase, they should look towards the target object more quickly in the Different Tones condition than in the Same Tones condition. Alternatively, if listeners cannot use the numeral's tone to predict the upcoming classifier and noun, then they should not be any faster in directing their eye gaze to the target object in the Different Tones condition compared to the Same Tones condition.

2.1 Methods

2.1.1 Participants

Participants were 43 right-handed, native Mandarin Chinese-speaking university students (37 female, age M = 22.8, $SD = 1.82$) studying in the UK. Prior to the experiment, participants spent on average 1.5 year overseas. All participants had normal or corrected-to-normal vision and normal hearing, and no neurological disorders. They all gave informed consent and were paid £7.5 per hour for their time.

2.1.2 Stimuli

The stimuli consisted of 44 sets of spoken sentences, each paired with a visual display of two objects. The sentences had non-constraining sentence frames and identified one of the two displayed objects at the end using a full noun phrase containing a numeral (either *yi* "one" or *liang* "two"), a classifier, and a head noun (the critical NP, e.g. *yi4 zhang1 chuang2*, one CL_{zhang} bed). The pair of objects on the visual display (see Figure 1) always required distinct classifiers, so that the classifier was equally informative about the target's identity across conditions. In the Different Tones condition, one of the object labels was associated with a classifier that would trigger a tone change in the numeral while the other was not, thus the two objects would require the numeral to be in different tones. As such, the tone of the numeral was highly informative about the identity of the upcoming classifier and noun. In the Same Tones (control) condition, neither of the object labels was associated with a classifier that would trigger a tone change in the numeral, so the tone of the numeral was completely uninformative about the identity of the target. If listeners can use the numeral's tone to anticipate the upcoming classifier and noun, they should be able to direct their eye gaze more quickly to the target object in the Different Tones condition than in the Same Tones condition. We tested the T3 sandhi using the numeral *liang* and the *yi* sandhi using the numeral *yi*, resulting in four conditions in total: *liang* Different Tones, *liang* Same Tones, *yi* Different Tones, *yi* Same Tones. A sample set of stimuli is shown in Figure 1.

The classifiers that were associated with the depicted objects were evenly distributed across the four tones. As such, objects associated with classifiers in each of the four tones were equally likely to be the target, (e.g. objects with a T1 classifier were the target 25% of the time). Since the tone sandhi used in the current experiment were triggered by one of the four tones (T4 in the *yi* sandhi, T3 in the T3 sandhi), the spoken numeral in the critical NP underwent tone changes in 25% of the experimental items. Picture positions were counter-balanced, such that the target picture was shown on the left side of the screen in half of the trials, and on the right in the other half.

Since the *yi* sandhi and the T3 sandhi are triggered in the critical NP by classifiers in two distinct tones, we constructed the stimuli such that objects used in the Same Tones condition with the numeral *liang* appeared in the Different Tones condition when the sentence had the numeral *yi*, and vice versa. An example of this can be seen in Figure 1.

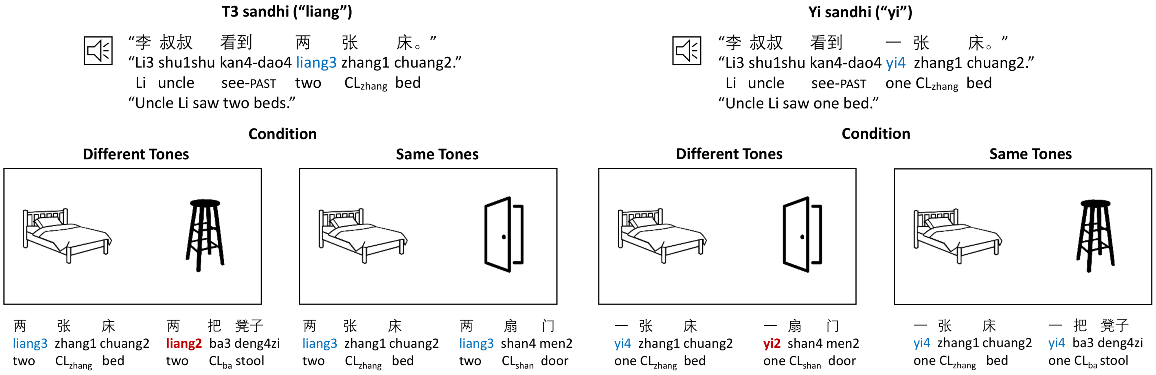

Figure 1: Illustration of stimuli used in Experiment 1. Unconstraining sentences ("uncle Li saw…") identified one of the objects in the visual display in a critical NP including a numeral, a classifier, and the target noun (*liang3 zhang1 chuang2*, two CL_{zhang} bed, "two beds"). In the Different Tones condition, the two objects' labels and their classifiers require the numeral to be in different tones as a result of tone sandhi applying to the numeral in one of them. As such, the tone of the numeral was informative about the identity of the upcoming classifier and noun. In the Same Tones (control) condition, neither classifier associated with the two object labels triggered a tone sandhi, so the tone of the numeral was uninformative. Visual displays that were used in the Different Tones condition for the T3 sandhi (*liang*) were used in the Same Tones condition for the *yi* sandhi, and vice versa.

2.1.2.1 Visual stimuli

Visual stimuli were black-and-white line drawings from the Noun Project (Polyakov et al., 2010) and the BCBL Multipic project (Duñabeitia et al., 2018). Images were selected via an identification task hosted on Gorilla Experiment Builder (Anwyl-Irvine et al., 2020), where ten native speakers of Mandarin Chinese were prompted to name the pictures by providing a classifier and a noun to fill in the sentence 图上展示的是一…… "The picture shows a …". A total of 81 images to which at least 70% of the participants responded with the same label and a classifier that required the numeral to be in the same tone were selected as visual stimuli. Finally, seven additionald images were included to reach a total of 88 images. The last seven images are from the tools-instrument class that commonly requires the classifier *ba3*[3].

2.1.2.2 Auditory stimuli Spoken sentences were recorded by a female native Mandarin speaker in a sound-proof booth. Intonation followed that of Mandarin Chinese declarative sentences. The recorder was instructed to say the sentences in a clear but natural manner. The average speech rate was 4.1 syllables per second ($sd = 0.54$), extracted using Praat (Boersma & Weenink, 2022).

On average, the duration of the syllable *liang* was 378ms, while *yi* had a duration of 272ms. Average pitch (F0) contours of the numerals (*liang* and *yi*) are presented in Figure 2.

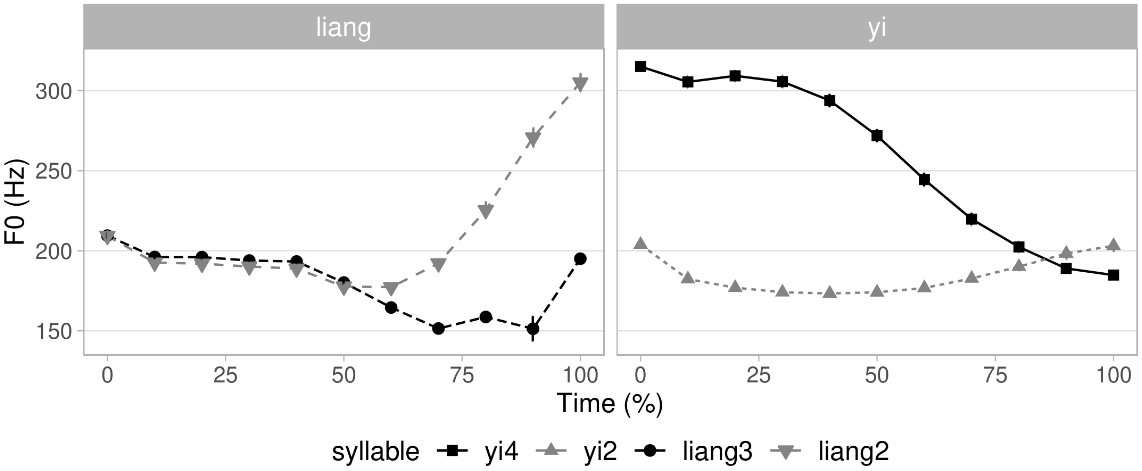

Figure 2: Average pitch (F0) contour of the numerals in the auditory stimuli of Experiment 1, within a normalized time window. Error bars represent ± 1 standard error of the mean. The average duration of *liang* was 378ms, and 272ms for *yi*.

2.1.2.3 Presentation lists and fillers Experimental items were arranged into four lists, such that each participant was only presented with one version of each item. Each participant thus saw 44 experimental items, half of which contained the numeral *liang* (T3 sandhi) and the other half *yi* (*yi* sandhi). Each list also contained 44 filler sentences, each paired with 2 black-and-white drawings of objects that were not in the experimental stimuli. Half (n=22) of the fillers had numerals that were neither *yi* nor *liang*, a quarter (n=11) had bare nouns (i.e., with no numerals or classifiers), and the rest (n=11) had the numeral *yi* or *liang*, but the two objects' labels shared the same classifier.

2.1.3 Procedure

Experiment 1 contained two parts: the main eye-tracking experiment (30min), as well as a classifier acceptability task (5min). The classifier acceptability task served to make sure that the participant used the same classifiers in their daily life as those used in the eye-tracking stimuli. For any classifier-noun pair that appeared in the eye-tracking stimuli, if the participant indicated that they would not use the pair in their daily life, trials containing that pair were subsequently removed from data analysis.

Participants were tested individually in a quiet room. The participant was seated 70 cm away from a 23" computer screen, and auditory stimuli were presented bilaterally via headphones. The eye-tracking procedure was programmed using E-Prime 2.0 (Schneider et al., 2012). Participant's eye movements were recorded using a Tobii TX300 eye-tracker with a sampling rate of 120Hz. At the beginning of each experimental session, a 9-point calibration procedure was performed.

At the beginning of each trial, a fixation cross appeared on the screen until eye gaze was detected on the cross, followed by the visual display of two objects. The spoken stimulus was played 1500ms after visual display onset. The images stayed on the screen until 100ms after audio offset. The order of trials was pseudo-randomized, with fillers and experimental items interleaved together. Following half of the fillers, participants were prompted to

[3] List of included unnormed nouns: *yuan2gui1* "bow compass", *qian2zi* "plier", *rui4shi4jun1dao1* "Swiss army knife", *chui2zi* "hammer", *deng4zi* "stool", *luo2si1dao1* "screw-driver", *da4ti2qin2* "cello".

answer a simple yes-no comprehension question about the sentence they have heard, by pressing one of two keys on the keyboard. The questions served to ensure that the participants attended to the stimuli throughout the experiment. Participants were allowed a break after every 22 trials. At the beginning of the experiment, participants completed two practice trials to familiarize themselves with the task. Figure 3 illustrates a typical trial in the eye-tracking experiment.

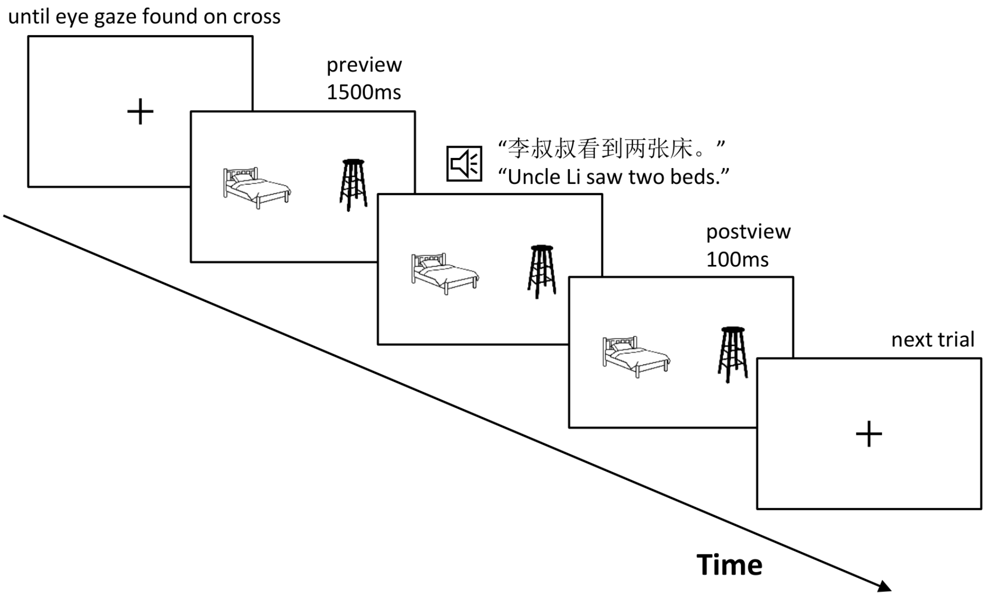

Figure 3: Eye-tracking procedure of Experiment 1. Once eye gaze was detected on the fixation cross, the visual display was presented for 1500ms before the onset of the spoken sentence. The visual display stayed present until 100ms after the offset of the spoken sentence.

Following the eye-tracking procedure, participants completed a classifier acceptability task. The task was hosted on Gorilla Experiment Builder (Anwyl-Irvine et al., 2020). At the beginning of the task, participants read a brief instruction explaining what classifiers were, with an example phrase. In each trial, participants saw a written noun phrase on the screen which consisted of the numeral/indefinite article *yi*, a classifier, and a noun (e.g., *yi2 ge4 ping2guo3*, one CL_{ge} apple, "an apple/one apple"). The participant was then asked to indicate if they found the classifier: (1) acceptable and they would use it with the given noun, (2) acceptable but that they are unlikely to use it themselves, or (3) unacceptable. Participants rated all 88 classifier-noun pairs that appeared in the eye-tracking experiment, in addition to 44 fillers in which the classifier was either one of the multiple classifiers that could be compatible with the noun, or incompatible with the noun.

2.1.4 Data analysis

Participants' comprehension accuracy was at ceiling level (accuracy $M = 0.98$, $SD = 0.04$), thus no one was excluded for low accuracy.

Data were processed in R (version 4.4.1, R Core Team, 2019). Based on the indivisual participants' responses in the classifier acceptability task, we excluded all traials where the participants indicated that they would not use the given classifier-noun pairing for one or both of the objects. (number of excluded trials = 122, 6.45% of all trials). In addition, a trial was also excluded if no eye gaze was detected on either object for over half the duration of that trial (excluded = 105, 5.55% of all trials). Finally, data from 1665 trials (88% of all trials) were included in the final analysis. Only the observations where the participant was fixating on one of the two objects were kept. The remaining data were then put into 20ms bins.

2.1.4.1 Divergence point analysis We adopted the divergence point analysis using a bootstrapping procedure to estimate the onset of looks to the target as proposed by Stone, Lago, et al. (2021). The procedure resamples the data and reapplies the estimation procedure multiple times, providing an average divergence point as well as its temporal variability for each condition. In our analysis, the divergence point is defined as the first time-point of 10 sequential time-bins where a one-sample t-test showed that looks to the target were significantly above chance (equivalent to the beginning of a 200ms window where there were significantly more looks to the target than chance). The result of the analysis is a distribution of onset times from resampling the data and reapplying the estimation 2000 times. During each resampling and reanalysis, the difference in onset between conditions was also calculated, resulting in a distribution of differences used to determine whether there is a significant difference between conditions, in terms of the onset of looks to the target object.

2.1.4.2 Applying Bayes' theorem to divergence points Following Stone (2021), we applied Bayesian principles to the distributions of divergence points, in order to assess the strength of evidence for and against the null hypothesis that listeners do not generate lexical predictions based on tone sandhi cues. This analysis uses Bayes' theorem to obtain a posterior distribution of divergence points and differences in divergence points. The posterior distribution is calculated from a prior distribution, representing the researcher's prior belief of the underlying effect, as well as a likelihood function, representing data observed in the experiment.

As the distributions of divergence points obtained from the bootstrapping procedure follow a normal distribution, we chose the normal distribution as our likelihood function, with means and standard distributions taken from the bootstrap data. A likelihood function was defined for each condition (T3 Different Tones, T3 Same Tones, *yi* Different Tones, *yi* Same Tones).

We defined principled prior distributions of divergence points: $\mathcal{N}(549, 174.5)$ for T3 sandhi trials, and $\mathcal{N}(515, 157.5)$ for *yi* sandhi trials, equivalent to normal distributions that allow the divergence point to fall within 200-899ms and 200-830ms after numeral onset 95% of the time for T3 sandhi trials and *yi* sandhi trials respectively. These time intervals represent a prior belief before observing data in the current experiment that the divergence points should fall within the interval between 200ms after numeral onset and 200ms after the average classifier offset, which was 699ms for T3 sandhi trials, and 630ms for *yi* sandhi trials. This belief is based on the principles that (1) sentential context was always unconstraining and gave no information about the identity of the target, thus the "point of disambiguation", when comprehenders identify the target, should not be earlier than the numeral onset; (2) native Mandarin Chinese listeners can use the classifier to ascertain the identity of the target (e.g., Huettig et al., 2010; Klein et al., 2012), thus the "point of disambiguation" should be no later than the classifier offset; and (3) roughly 200ms is needed to initiate a saccadic eye movement, thus the eye movement divergence point should be approximately 200ms after the "point of disambiguation".

Following the choice of priors, the likelihood functions were defined based on the data observed in the current experiment, namely the distributions of divergence points obtained in the divergence point analysis. Finally, posterior distributions of divergence points were determined for each condition, using Bayes' theorem: the posterior distribution is the product of the prior distribution and the likelihood.

To calculate Bayes factors as an index of strength of evidence, we determined prior and posterior distributions of the *difference* in divergence points between the Same Tones and Different Tones conditions. The prior distribution of difference is defined as a normal distribution, whose mean is the difference between the mean of the prior distributions in the Different Tones condition and in the Same Tones condition; this is zero since we used the same prior distributions for the two conditions, reflecting the prior belief that they were not different. The standard deviation of the prior distribution of difference is calculated as the square root of the sum of the squares of the prior standard deviation in the Different Tones condition and the prior standard deviation in the Same Tones condition. Posterior distributions of difference were defined in the same way using posterior means and standard deviations.

Bayes factor BF10, which indicates the strength of evidence in favour of $H1$ over $H0$, was determined for both T3 sandhi trials and *yi* sandhi trials using the Savage-Dickey equation ($BF10 = \frac{\text{prior density at null hypothesis}}{\text{posterior density at null hypothesis}}$). A BF10 above 1 provides evidence for $H1$ ($1 <$ weak evidence for H1 $< 3 <$ moderate evidence $< 10 <$ strong evidence $< +\infty$), while a BF10 below 1 provides evidence for $H0$ ($0 <$ strong evidence for H0 $< \frac{1}{10} <$ moderate evidence $< \frac{1}{3} <$ weak evidence < 1).

2.2 Results

2.2.1 Divergence point results

Figure 4 shows the proportion of looks to the target and competitor objects. In the T3 sandhi trials, the estimated divergence point was 658ms following numeral onset (95% CI = [580, 740]) in the Different Tones condition, and 746ms [700, 800] in the Same Tones condition. The divergence point appeared 88ms earlier in the Different Tones condition than the Same Tones condition, with a 95% CI of [-180, 0], covering negative values and zero. This suggests a tendency towards a significant difference between the onset of looks to the target object in the Different Tones condition vs. in the Same Tones condition.

In the *yi* sandhi trials, the estimated divergence point was 609ms [560, 680] in the Different Tones condition, and 621ms [580, 660] in the Same Tones condition. The divergence point was only 12ms earlier in the Different Tones condition than the Same Tones condition with a 95% CI of [-60, 80], covering both negative and positive values. Figure 5 shows the distribution of divergence point differences.

2.2.2 Applying Bayes' theorem to divergence points

Applying Bayes' theorem to divergence points revealed posterior distributions of divergence points for each condition: T3 sandhi Different Tones $\mathcal{N}(653, 38)$, T3 sandhi Same Tones $\mathcal{N}(743, 24)$; *yi* sandhi Different Tones $\mathcal{N}(607, 27)$, *yi* sandhi Same Tones $\mathcal{N}(619, 24)$. The posterior *difference* between conditions were $\mathcal{N}(-90, 45)$ for T3 sandhi trials and $\mathcal{N}(-12, 36)$ for *yi* sandhi trials (Figure 6).

A Bayes factor $BF10 = 1.33$ for T3 sandhi trials provide a weak support for the $H1$ that listeners used T3 sandhi cues to generate lexical predictions. Meanwhile, a Bayes factor $BF10 = 0.17$ for *yi* sandhi trials provide moderate evidence for the $H0$ that listeners did *not* use *yi* sandhi cues to generate lexical predictions.

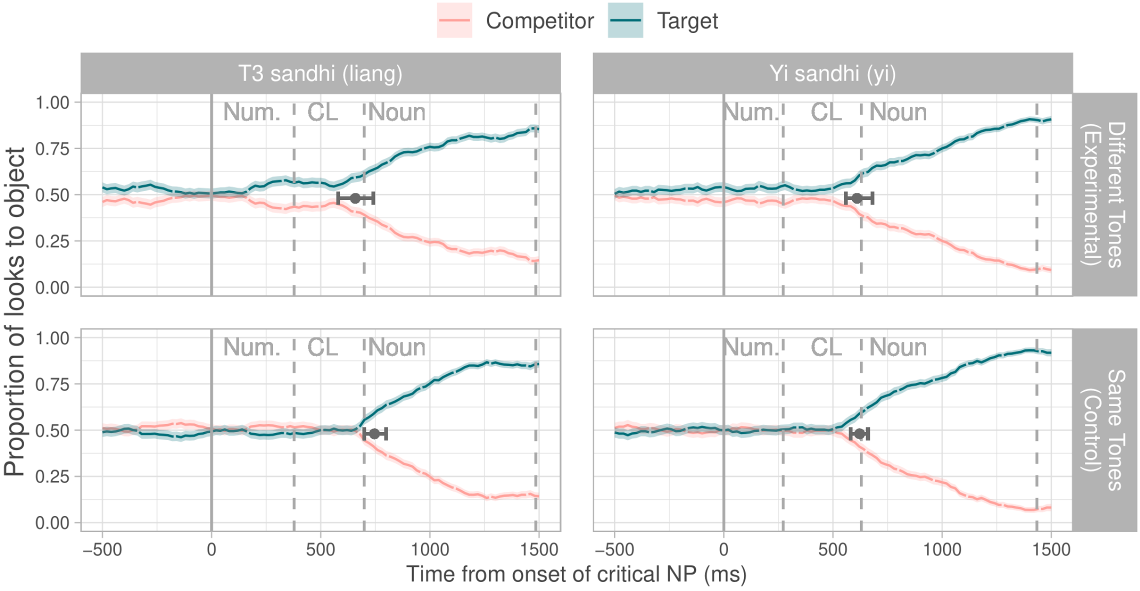

Figure 4: Proportion of looks to the target and competitor objects, Experiment 1 (Left column: T3 Sandhi; Right column: Yi Sandhi; Top row: Different Tones condition; Bottom row: Same Tones condition). The solid vertical line shows the onset of the critical NP (i.e. the onset of the numeral). The dashed lines show the average onset of the classifier and noun, as well as the offset of the noun. The solid point represents the mean divergence point estimated from the bootstrapping procedure, the error bars represent the 95% confidence interval.

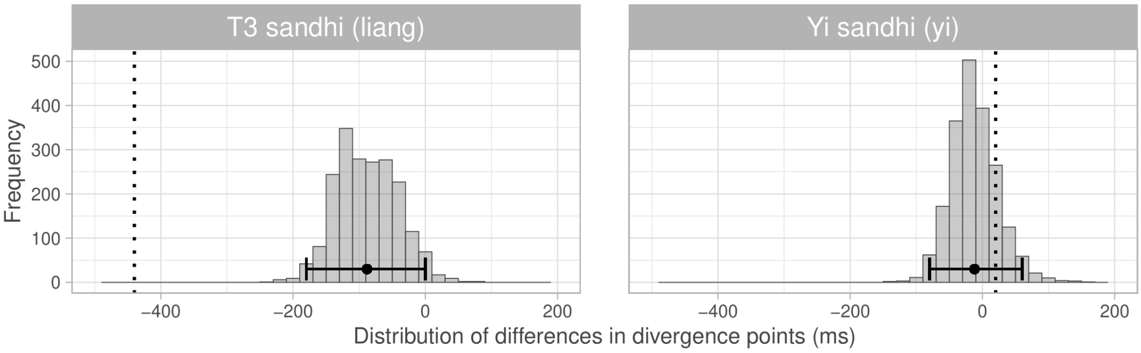

Figure 5: Bootstrapped distribution of differences in divergence points between the conditions (Different Tones minus Same Tones) in the T3 sandhi (left) and *yi* sandhi trials (right), Experiment 1. The solid point and the error bars represent the mean difference and the 95% confidence interval respectively. The vertical dotted lines represent the difference calculated from doing the estimation procedure on the original data before resampling.

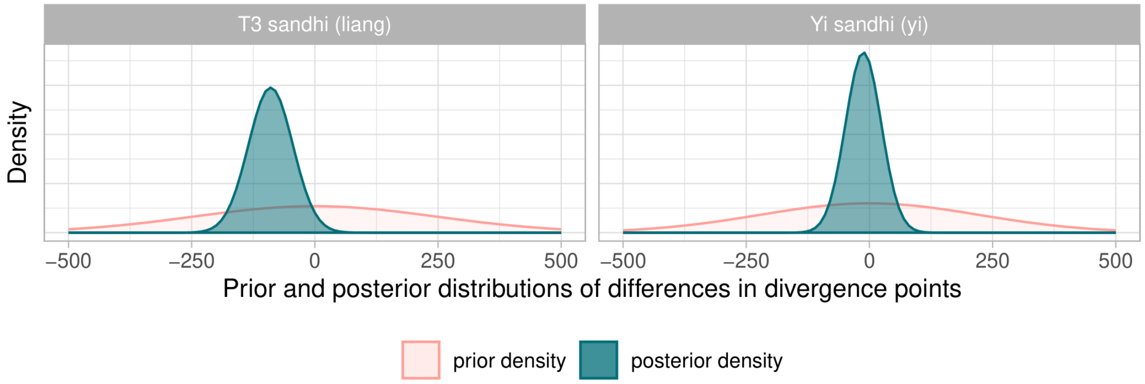

Figure 6: Prior and posterior distributions of differences in divergence points in the T3 sandhi (left) and *yi* sandhi trials (right), Experiment 1.

2.2.3 An exploratory analysis of base vs. sandhi tones in T3 sandhi trials

As we observed a tendency towards a significant prediction effect for T3 sandhi trials, we further explored the T3 sandhi data by separating trials where listeners heard the base tone *liang3* from those where they heard *liang2* in the Different Tones condition.

The auditory stimuli contained the sandhi tone *liang2* in the Different Tones condition for half of the experimental items (where the object that triggered a T3 sandhi was named), and the base tone *liang3* for the other half (where the object that did not trigger a T3 sandhi was named). All items contained the base tone *liang3* in the Same Tones condition. We separately analysed experimental items where a sandhi tone *liang2* was heard in the Different Tones while a base tone *liang3* was heard in the Same Tone condition, versus experimental items where a base tone *liang3* was heard in both conditions (Figures 7, 8).

For items with a base tone *liang3* in both conditions, divergence point analyses revealed a mean onset of looks to the target object of 760ms [680, 840] in the Different Tones condition and 773ms [700, 860] in the Same Tones condition. A difference (Different Tones minus Same Tones) of -12ms [-140, 100] indicates no evidence for prediction based on the T3 sandhi when a base tone was heard in the both conditions.

For items that contained a sandhi tone *liang2* in the Different Tones condition, the mean onset of looks to the target object was 720ms [640, 820] in the Different Tones condition and 833ms [780, 900] in the Same Tones condition. A difference of -112ms [-200, -20] indicates that listeners directed their eye gaze towards the target object significantly earlier in the Different Tones condition than in the Same Tones condition when they heard a sandhi tone in the Different Tones condition.

In sum, the exploratory analysis of base vs. sandhi tones in T3 sandhi trials reveals that listeners were able to compute a lexical prediction based on the sandhi tone *liang2*, but not based on the base tone *liang3*.

2.3 Discussion

In Experiment 1, we used the T3 sandhi and *yi* sandhi in Mandarin Chinese noun phrases as a test case to investigate whether listeners use phonological information as an input to lexical prediction during language comprehension. In a visual world eye-tracking experiment, listeners heard sentences with a non-constraining context that identified one of the two objects on the visual display with an NP that contained a numeral followed by a classifier and a noun. The tone of the numeral was informative about the target's identity in the Different Tones condition but uninformative in the Same Tones condition. Results showed that listeners tended to direct their eye gaze towards the target object more quickly in the Different Tones condition than in the Same Tones condition in the T3 sandhi trials, but not in the *yi* sandhi trials. Further, our supplementary analysis revealed that an effect of prediction was observed only when the numeral has undergone tone sandhi and was realised in its sandhi tone *liang2*, but not when it was in its base tone *liang3*, suggesting that the predictive effect of T3 sandhi may have been driven by trials where the numeral was realised in its sandhi form.

These initial results suggest that listeners of Mandarin Chinese can use the T3 sandhi in the numeral *liang* to generate predictions about an upcoming classifier and noun, especially when it is realised the sandhi tone. Meanwhile, however, we found no evidence that the listeners could use the *yi* sandhi in the numeral to predict the upcoming classifier and noun.

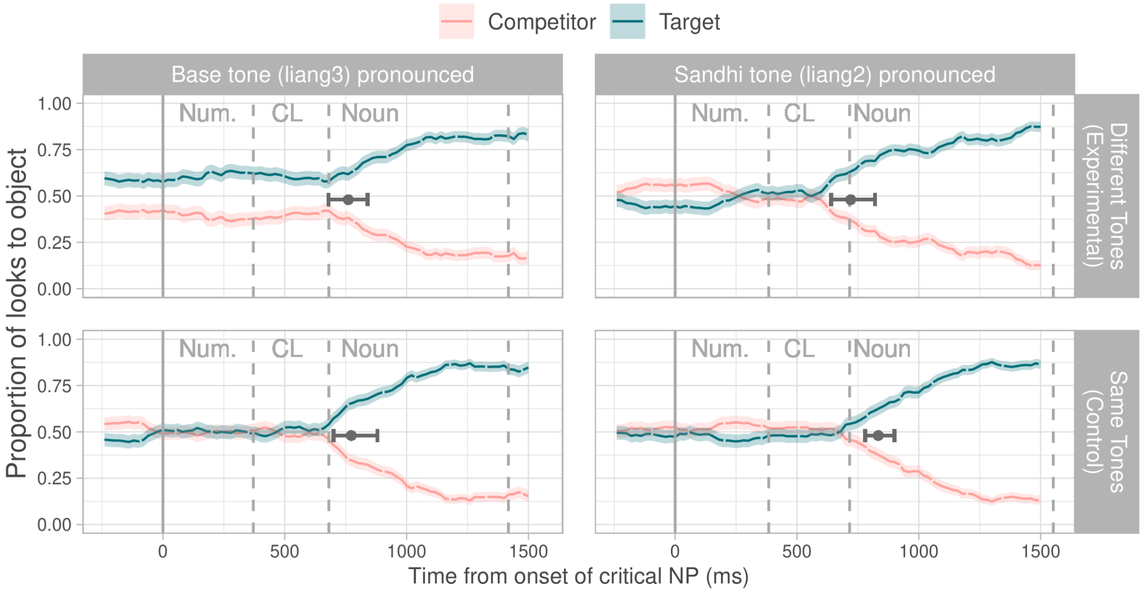

Figure 7: Proportion of looks to the target and competitor objects, Experiment 1, T3 sandhi trials. Left column: trials with *liang3* in both conditions; Right column: trials with *liang2* in the Different Tones condition and *liang3* in the Same Tones condition; Top row: Different Tones condition; Bottom row: Same Tones condition.

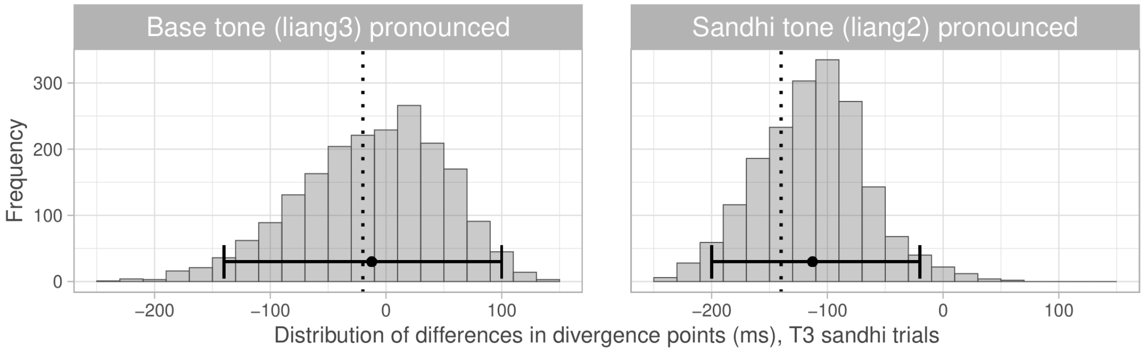

Figure 8: Bootstrapped distribution of differences in divergence points between the conditions (Different Tones minus Same Tones), Experiment 1, T3 sandhi trials. Left: trials with *liang3* in both conditions; Right: trials with *liang2* in the Different Tones condition and *liang3* in the Same Tones condition.

The lack of a predictive effect in the *yi* sandhi trials contrasts with the results reported by Liu et al. (2025). Liu et al. (2025) had a similar experimental manipulation and found that Mandarin Chinese listeners were able to use the tone of *yi* to look anticipatorily at the target object, while we found no such evidence. We believe the different results may be attributable to two key differences between the two studies. Firstly, while the present study looked at both the numerals *yi* and *liang*, Liu et al. (2025) focused exclusively on the numeral *yi.* Overall, the numeral *yi* appeared in 32% of our stimuli and 100% of Liu et al. (2025)'s as they did not include any filler items. As a result, listeners were likely able to anticipate the presence of *yi* in all trials in Liu et al. (2025)'s study.

Secondly, while sentences were produced naturally at a normal speech rate of 4.1 syllables per second in the present experiment, Liu et al. (2025) presented their sentences syllable-by-syllable at an artificially slow, fixed speech rate (800ms per syllable). As a result, *yi* cues might have been highly salient in Liu et al. (2025)'s study, making participants more likely to become aware of their informativeness. Considering the differences in experimental designs, the present experiment suggests that listeners of Mandarin Chinese may not regularly use *yi* sandhi cues to generate lexical predictions.

Interestingly, the observation of a predictive effect in *liang* trials but not in *yi* trials in the present experiment suggests that listeners may be better at using the T3 sandhi than the *yi* sandhi in numerals to predict the upcoming classifier and noun. This may be explained by some differences between the T3 sandhi and the *yi* sandhi.

In particular, the T3 sandhi applies to all T3 syllables in Mandarin, whereas the *yi* sandhi only applies to the specific word or morpheme *yi* (the numeral "one"). Although *yi* is a highly frequent morpheme in Mandarin (5895 per million according to SUBTLEX-CH), the overall frequency of all words in which the T3 sandhi applies is much higher (14562 per million) (Cai & Brysbaert, 2010). In addition, since T3 sandhi can apply across word boundaries, the real difference in frequency between the two sandhi patterns is likely even greater than these raw word frequencies suggest.

This means that native speakers of Mandarin Chinese are likely to have considerably more experience with the T3 sandhi than with the *yi* sandhi in their everyday language use. Consequently, comprehenders may find it easier to use T3 sandhi than the *yi* sandhi as a cue for predicting upcoming language input because they encounter the T3 sandhi more frequently and are more practiced in processing it compared to the *yi* sandhi.

3 Experiment 2

Results from Experiment 1 suggest that listeners were able to use the T3 sandhi pattern in numerals to generate predictions during language comprehension, especially when the prediction cue was the sandhi form of a T3 syllable. Meanwhile, we found no evidence that listeners could use the *yi* sandhi to generate predictions. A possible explanation for the difference between the two sandhi patterns is the potential different levels of sensitivity to them.

In the present experiment, therefore, we used an acceptability rating task to investigate to what extent listeners are sensitive to violations of the *yi* sandhi as well as the T3 sandhi. Evidence from lexical identification experiments and ERP studies shows that native speakers of Mandarin Chinese are very sensitive to lexical tone contrasts in general (Hallé et al., 2004; Xi et al., 2010), but it is less clear to what extent listeners are sensitive to tone sandhi patterns and their violations: listeners may know that *yi2* and *yi4* are tonally distinct, but they may treat them as allophonic in a phrase involving the numeral *yi* if they are not sensitive to the *yi* sandhi's violations.

Participants were presented with written disyllabic (two-character) phrases in Mandarin Chinese that either involved the T3 sandhi or the *yi* sandhi together with audio recordings that were either consistent with the relevant tone sandhi pattern (the Grammatical condition) or violated it (the Ungrammatical condition) (e.g. *yi2 ge4* vs. **yi4 ge4*). We distinguished tone sandhi violation in two different phonological contexts: sandhi-inducing contexts and non-sandhi-inducing contexts. Tone sandhi violation in a sandhi-inducing context involved a syllable not undergoing tone change when change was required by a neighbouring syllable (e.g. T3 remaining low-dipping in T3T3). Meanwhile, in a non-sandhi-inducing context, tone sandhi violations occurred when a syllable underwent unwarranted tone change (e.g. T3 realised as a rising T2 when the following syllable was T1/T2/T4).

For the *yi* sandhi, we referred to *yi*-T4 as the sandhi-inducing context and *yi*-T1/T2/T3 as the non-sandhi-inducing context. This was because (i) the numeral *yi* was never pronounced in isolation in the experiment, so the true base tone of *yi* (*yi1*) was never heard, and (ii) *yi* is realised in T2 only when followed by a T4 and in T4 when followed by a syllable in the other three tones. We measured listeners' sensitivity to tone sandhi

	sandhi-inducing context	non-sandhi-inducing context
Grammatical, *liang*	两本 / *liang2 ben3*	两个 / *liang3 ge4*
Ungrammatical, *liang*	两本 / **liang3 ben3*	两个 / **liang2 ge4*
Grammatical, *yi*	一个 / *yi2 ge4*	一本 / *yi4 ben3*
Ungrammatical, *yi*	一个 / **yi4 ge4*	一本 / **yi2 ben3*

Table 1: An example set of stimuli used in the acceptability rating task. Participants saw written Chinese characters on the screen (e.g. 两本, *liang2 ben3*) while listened to a pronunciation that either was consistent with the relevant tone sandhi (the Grammatical condition) or violated the tone sandhi (the Ungrammatical condition). Tone-sandhi-violating pronunciations are marked with an asterisk.

violations by comparing their acceptability ratings between the Grammatical and Ungrammatical conditions. If listeners are less sensitive to the *yi* sandhi than the T3 sandhi, we would expect an interaction between Sandhi and Grammaticality, such that the effect of Grammaticality would be greater in the T3 sandhi than the *yi* sandhi items.

3.1 Methods

3.1.1 Participants

Participants were 40 native Mandarin speakers living in mainland China (32 female and 8 male, age M = 22.6, SD = 2.1). All participants had normal hearing, normal or corrected-to-normal vision, and no known neurological disorders. No participants in this experiment also participated in Experiment 1.

3.1.2 Stimuli

Stimuli were 48 disyllabic phrases composed of a numeral (T3 sandhi: *liang* or *yi* sandhi: *yi*) and a classifier. In the Ungrammatical condition, the tones of the numeral and the classifier violated the tone sandhi patterns (e.g. **liang3 ben3*); the tones of these syllables were consistent with the tone sandhi patterns (e.g. *liang2 ben3*) in the Grammatical condition. In addition, we manipulated phonological context (sandhi-inducing vs. sandhi-non-inducing). In the sandhi-inducing context, the second syllable is in a tone that should trigger tone sandhi in the first syllable; here, numerals were in their sandhi tones (*liang2 ben3*, *yi2 ge4*) in the Grammatical condition, and in their base tones (**liang3 ben3*, **yi4 ge4*) in the Ungrammatical condition. In the non-sandhi-inducing context, the second syllable normally does not trigger tone sandhi; in this context, the numerals' base tones were in the Grammatical conditions (*liang3 ge4*, *yi4 ben3*), and sandhi tones were in the Ungrammatical conditions (**liang2 ge4*, **yi2 ben4*). A total of 48 classifiers (24 in T3, 24 in T4) were used. Stimuli were divided into four lists, such that each item would only appear in a list once. A sample set of stimuli is shown in Table 1.

Fillers consisted of 48 disyllabic phrases or words, of which 16 contained a numeral other than *yi* or *liang* and a classifier, 32 were disyllabic nouns or verbs and did not have any numeral or classifier (e.g. *ji4jie2* "season", *xia4zai4* "download"). Half of the fillers were correctly pronounced, a quarter of the fillers had an incorrect tone in one of the syllables (e.g. *zhao4pian1* "photo": **zhao4pian4*), while the remaining quarter had a segment error (an error created by substituting one of the segments) (e.g. *xue2ke1* "school subjects": **xue2ge1*). Half of the errors occurred on the first syllable, the other half on the second syllable. No errors in the fillers involved either the T3 sandhi or the *yi* sandhi.

A female native Mandarin speaker recorded the auditory stimuli. All stimuli, including practice and filler items, were spliced using tokens from two separate, natural recordings where no tone sandhi rules were violated. For example, for the stimulus **liang3 ben3* (which is ungrammatical in a T3 sandhi-inducing context), the two syllables were spliced together from the recordings of "*liang3 ben4*" and "*liang2 ben3*" respectively; for the stimulus *liang2 ben3* (which is grammatical in a T3 sandhi-inducing context), the two syllables were from two different recordings of "*liang2 ben3*".

3.1.3 Procedure

The experiment was conducted online using Gorilla Experiment Builder (Anwyl-Irvine et al., 2020). Participants completed the experiment on their computers, tablets, or smartphones. Participants were instructed to use headphones, and to start the experiment only when they were in a quiet environment.

At the beginning of each trial, a fixation cross appeared in the middle of the screen for 300ms. Then, the two-character, disyllabic Chinese phrase was visually presented at the centre of the screen, while the auditory stimulus was played twice, with a 1000ms silent interval in between. Participants were asked to rate how natural the recording sounded as a pronunciation of the target phrase shown on the screen, on a seven-point scale (1 = totally unnatural to 7 = totally natural). Participants were instructed to rate the recording as "unnatural" if either of the two syllables sounded unnatural. Participants completed four practice trials before the start of the experiment. The practice stimuli consisted of one numeral-noun phrase (*yi4 tian1* "one day"), and three disyllabic words. Among the four practice trials, one had a tone error, and another had a segment error. Feedback was given on the screen after each practice trial to encourage participants to use the full range of the scale. The participant was prompted to take a break after every 40 trials.

After completing the acceptability rating task, the participant was given a short questionnaire about their knowledge of the T3 sandhi and the *yi* sandhi. A short definition of the T3 sandhi and the *yi* sandhi was shown on the screen, and participants were asked whether they were aware of the patterns of the T3 sandhi as well as the *yi* sandhi prior to reading the questionnaire, and whether they had been formally taught those rules in school.

3.1.4 Data analysis

Responses were z-transformed using by-participant means. Z-transformed responses were analysed with linear mixed-effects models (Baayen et al., 2008) with the lme4 package (Bates et al., 2015), the emmeans package (Lenth, 2024), and the cAIC4 package (Säfken et al., 2021) for R 4.4.1 (R Core Team, 2019). All categorical variables were sum contrast coded (sandhi: *yi* sandhi = -1, T3 sandhi = 1; context: inducing = -1, non-inducing = 1; grammaticality: grammatical = -1, ungrammatical = 1). Linear mixed-effects models[4] were fitted to the data, with fixed effects of sandhi (*yi* vs. T3), context (sandhi inducing vs. non-sandhi-inducing), and grammaticality (Grammatical vs. Ungrammatical), and interactions among the fixed effects. For random effects structure, we started from a maximal structure, with random intercepts and random slopes for both participant and item (Barr et al., 2013). As the maximal model failed to converge, we simplified the model by only including random by-participant and by-item intercepts as random effects. Where interactions were found, mixed-effects linear models were fitted to subsets of data to investigate the interactions.

3.2 Results

3.2.1 Language experience

Participants' awareness and their reported classroom experience did not differ significantly between the T3 sandhi and the *yi* sandhi (Table 2, classroom experience: $X^2(4) = 5.43$, $p = 0.25$; awareness: $X^2(2) = 2.4$, $p = 0.3$).

	the T3 sandhi			the *yi* sandhi		
	yes	no	do not remember	yes	no	do not remember
explicit knowledge	25 (62.5%)	15 (37.5%)		29 (72.5%)	11 (27.5%)	
classroom experience	18 (45%)	10 (25%)	12 (30%)	19 (47.5%)	4 (10%)	17 (42.5%)

Table 2: Responses (percentage of answers) to the post-experiment questionnaire in Experiment 2. Explicit knowledge was assessed using the question "你做问卷前是否知道一字变调/三声变调的知识？ " (Did you know about the *yi* sandhi/the T3 sandhi before completing the questionnaire?). Classroom experience was assessed using the question "请回忆你在语文课上的学习，老师是否在课上教过一字变调/三声变调的知识？ " (Please recall your Chinese classes, did your teacher ever talk about the *yi* sandhi/the T3 sandhi?). The "do not remember" option was only available for the classroom experience questions.

3.2.2 Linear mixed-effects models

Figure 9 shows participants' z-transformed scores across conditions. The mixed-effects linear model revealed main effects of sandhi ($\beta = 0.18$, $t = 20.06$, $p < .001$), context ($\beta = -0.17$, $t = -13.57$, $p < .001$), as well as

[4] The mixed-effect linear model: lmer(z ~ sandhi*context*grammaticality + (1|ParticipantID) + (1|item), REML = FALSE).

grammaticality ($\beta = -0.86$, $t = -67.32$, $p < .001$). All the 2-way and 3-way interactions were also significant.

To follow up the three-way interaction, two linear mixed effect models with fixed effects of grammaticality and context were fitted to the data from the T3 sandhi and the *yi* sandhi trials, respectively.

For the T3 sandhi trials, both predictors showed main effects (context: $\beta = -0.36$, $t = -18.99$, $p < 0.001$; grammaticality: $\beta = -0.69$, $t = -45.04$, $p < 0.001$). There was also a significant interaction between grammaticality and context ($\beta = -0.36$, $t = -23.56$, $p < 0.001$). Conditional Akaike information criterion (cAIC) confirmed that the model including both predictors and their interaction best explained the data ($cAIC = 1309.24$; model without interaction: $cAIC = 1742.2$; model without context: $cAIC = 1783.81$). Analysis of data in the sandhi-inducing context revealed a significant effect of grammaticality ($\beta = -0.65$, $t = -15.17$, $p < 0.001$). The same effect was found in the non-sandhi-inducing context, as well, although the effect size is much larger ($\beta = -2.09$, $t = -48.45$, $p < 0.001$).

For the *yi* sandhi trials, there was only a significant main effects of grammaticality ($\beta = -1.04$, $t = -105.74$, $p < 0.001$), and no interaction between context and grammaticality.

Notably, we found a significant main effect of grammaticality overall and in all subsets of data, such that grammatical trials were rated significantly more acceptable than ungrammatical trials. Moreover, phonological context played a role in listeners' acceptability ratings of T3 sandhi violations, such that the effect of grammaticality was smaller in a sandhi-inducing context than a non-sandhi-inducing contexts.

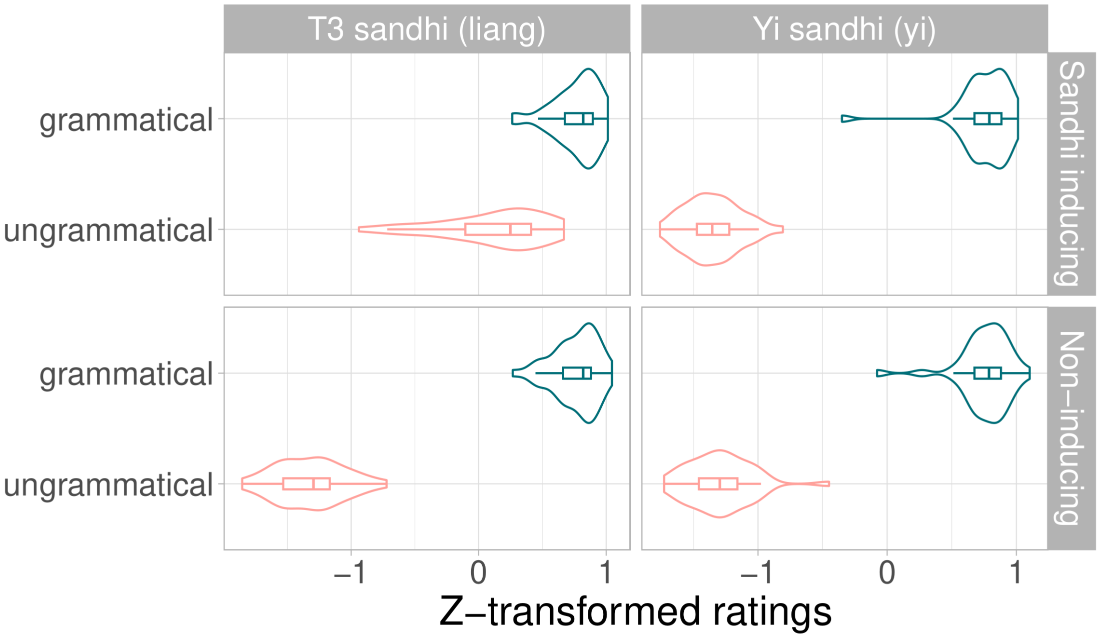

Figure 9: Z-transformed acceptability ratings by condition, aggregated by subject, Experiment 2.

3.3 Discussion

The current experiment sought to investigate a potential explanation for the results of Experiment 1. In Experiment 1, we found that informative T3 sandhi cues in the numeral *liang* resulted in anticipatory eye movements to the target object, but informative *yi* sandhi cues did not. We hypothesised that this may be because listeners may be less familiar with the *yi* sandhi and thus be less sensitive to the constraints that the *yi* sandhi poses on the following syllable than the T3 sandhi. The current experiment tested this possibility using an acceptability judgment task. Listeners saw written numeral-classifier phrases on the screen while they listened to a recording of those phrases, which either was consistent with or violated the relevant tone sandhi. Responses to the tone sandhi experience questionnaire did not suggest any significant differences between the T3 sandhi and the *yi* sandhi in the proportion of participants who were aware of the tone sandhi patterns and who were formally instructed about them. Acceptability ratings suggested that listeners were highly sensitive to violations of both sandhi patterns (a significant main effect of grammaticality). Crucially, listeners' sensitivity to the *yi* sandhi was not any lower than that to the T3 sandhi, suggesting that listeners of Mandarin Chinese have robust knowledge of both sandhi patterns.

Taken together, the results from the current experiment and Experiment 1 thus suggest that although listeners have robust knowledge of the *yi* sandhi and the constraints it poses on upcoming language input, they may not effectively use it to generate predictions during real-time language comprehension. This indicates that listeners may not always use all informative phonological cues they are sensitive to in computing predictions about upcoming words. This is consistent with A. Ito & Hirose (2024)'s findings that although Japanese listeners used informative low tones to predict, they did not do so with high tones that were equally informative.

The current experiment demonstrated that listeners were familiar with the constraints that *yi* sandhi poses on the following syllable, thus it cannot explain the lack of evidence for prediction based on the *yi* sandhi in Experiment 1. However, although the difference between language experiences of the *yi* sandhi and the T3 sandhi did not result in a difference in their (offline) knowledge, there may still be differences between how these two sandhi patterns are used in real-time language comprehension. For example, listeners may be slower to use a *yi* sandhi cue to generate lexical predictions than a T3 sandhi cue, such that effects of prediction with the *yi* sandhi can only be observed when the listener has plenty of time. We shall return to the issue of the time course of prediction in the General Discussion.

3.3.1 Reduced sensitivity to T3 sandhi violations

In addition to the observation that listeners are highly sensitive to the *yi* sandhi, we also found a reduced sensitivity to T3 sandhi violations in the sandhi-inducing context (T3T3). This is consistent with previous findings on T3 sandhi perception. Li & Chen (2015) measured native listeners' mismatch negativity (MMN) to tonal deviants in a passive oddball paradigm in an electroencephalography (EEG) experiment. Participants listened to repeated presentations of a syllable (*ma*) in either T1, T2, or T3, with occasional presentations of a deviant syllable that differed from the standard syllable in terms of lexical tone. Brain responses showed reduced MMN response to T2 deviants in T3 standards compared with the reverse (T3 deviants in T2 standards) or other tonal pairs that do not involve the T3 sandhi (T1/T3 deviants in T3/T1 standards). The authors took this to suggest that native speakers' mental representation of the third tone may include its sandhi form, which is highly similar to the T2. Thus, the perceptual difference of a T2 deviant from T3 standards is reduced, as their representation of the sandhi form T2 is also activated when listening to T3 standards.

Similarly, A. Chen et al. (2015) used a differentiation task to investigate the perception of T3 sandhi. Participants listened to pairs of disyllabic phrases and were asked to decide whether the two phrases were the same or different. A. Chen et al. (2015) tested two types of combinations, T2T3 - T3T3 pairs (i.e. grammatical vs. ungrammatical pairs), and T1T4 - T4T4 pairs as control. Results suggested that L1 speakers of Mandarin Chinese had more difficulty distinguishing T2T3 phrases from T3T3 phrases than distinguishing T1T4 phrases from T4T4 phrases. These results can also be explained by a representation of the T3 sandhi in which hearing the base form T3 also activates the sandhi form T2 to an extent.

According to this view, in the current experiment, upon hearing the first T3 syllable, the sandhi tone T2 may become partially activated, thus the sandhi-violating T3T3 sequence may be perceived as more acceptable than other cases of sandhi violations. Nevertheless, there was a significant difference between the acceptability ratings given to a sandhi-consistent T2T3 phrase vs. a sandhi-violating T3T3 phrase, indicating that listeners were still able to distinguish between an ungrammatical T3T3 phrase from a grammatical T2T3 phrase most of the time.

This reduced sensitivity to T3T3 may also explain why prediction effect was found in trials that involved a sandhi tone *liang2* (but not the base tone *liang3*) in Experiment 1. For example, in the visual display *liang3 zhang1 chuang2* (two CL$_{\text{zhang}}$ beds) vs. *liang2 ba3 deng4zi* (two CL$_{\text{ba}}$ stools), upon hearing *liang3* in the spoken sentence, if listeners were happy to accept a T3 syllable following *liang3* to a certain extent, then both objects in the visual display would be compatible with the auditory input, rendering *liang3* uninformative about the identity of the target.

3.3.2 A large variability in listeners' sensitivity to T3T3

In addition to the reduced sensitivity to T3T3 violations, we further observed a wider distribution of acceptability ratings given to this type of tone sandhi violation than the other violations. Figure 10 shows the by-participant mean ratings to the ungrammatical trials. The top left panel shows that participants' ratings to T3 sandhi violations in the sandhi-inducing context is subject to an increased between- and within-participants variability compared to other types of sandhi violations.

The large variability in the acceptability scores suggest that there is a larger variability in the sensitivity to T3T3 sequences, both within- and between-individuals. For the current study, this suggests that the extent to which a listener can use a T3 sandhi cue to predict may also be subject to a large variability in Experiment

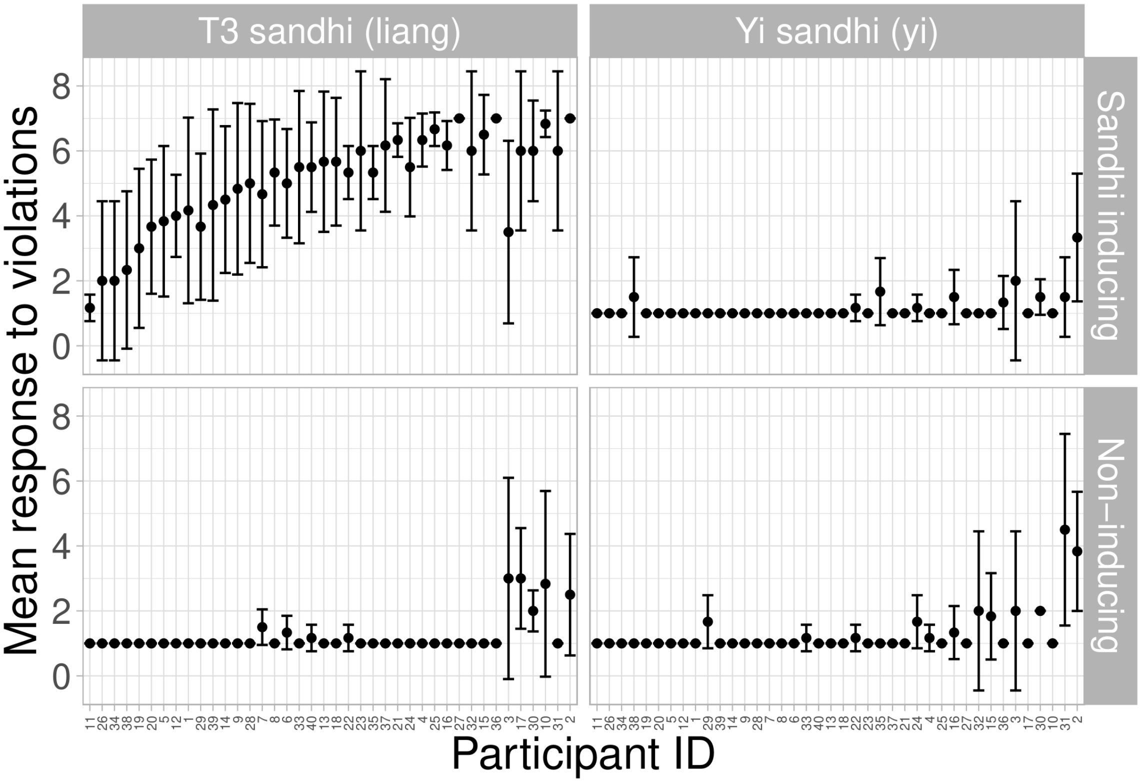

Figure 10: By-subject mean ratings given to ungrammatical trials, ordered by average ratings given to ungrammatical trials. Error bars show standard deviations.

1: for listeners who were more sensitive to T3T3 sequences, both objects on the visual display may be highly compatible with the auditory input *liang3*, effectively rendering it uninformative about the target; whereas for listeners who were more sensitive to this type of violations, only the target object will be compatible with the auditory input, and T3 sandhi cues can be highly informative about the target's identity.

Therefore, in Experiment 3, we collected eye-tracking and acceptability rating data from a single group of participants, to investigate whether listeners' sensitivity to the T3 sandhi violations in a sandhi-inducing context is correlated with the extent to which they can use the T3 sandhi in prediction during spoken language comprehension.

4 Experiment 3

Results in Experiment 1 suggested that native Mandarin Chinese listeners used the T3 sandhi in the numeral *liang* to predict an upcoming classifier and noun, especially when the sandhi tone *liang2* served as the prediction cue. Meanwhile, there was no evidence that showed listeners could use the *yi* sandhi in prediction. Experiment 2 showed that listeners were no less sensitive to violations of the *yi* sandhi than those of the T3 sandhi, but there was a large variability in the acceptability of violations of the T3 sandhi in the sandhi-inducing context (T3T3). In the present experiment, we collect eye-tracking data and tone sandhi sensitivity data from a single group of participants to investigate whether listeners' ability to use the T3 sandhi to make predictions during real-time comprehension correlates with their sensitivity to the T3 sandhi.

4.1 Methods

4.1.1 Participants

Participants were 43 right-handed, native Mandarin-speaking university students (40 female, age M = 20.7, SD = 2.26) studying in the UK (average life spent overseas = 1.2 years). All participants had normal or corrected-to-normal eyesight and normal hearing, and no neurological disorders. They all gave informed consent and were paid £7.5 per hour for participating.

4.1.2 Stimuli

Stimuli for the eye-tracking part as well as the classifier acceptability task were the same as Experiment 1. Stimuli for the tone sandhi sensitivity task were the same as Experiment 2.

4.1.3 Procedure

Experiment 3 contained three parts: the main eye-tracking experiment (30min), the classifier acceptability task (5min), as well as the tone sandhi sensitivity task (15min), in chronological order. The procedure of the first two parts was identical to Experiment 1 and that of the third part directly replicated Experiment 2.

4.1.4 Data analysis

4.1.4.1 Tone sandhi sensitivity Linear mixed-effects models with fixed effects of sandhi (T3 vs. *yi*), context (sandhi-inducing vs. non-sandhi-inducing), and grammaticality (Grammatical vs. Ungrammatical), as well as interactions among the fixed effects were fitted to z-transformed acceptability ratings with the lme4 (Bates et al., 2015), emmeans (Lenth, 2024), and the cAIC4 (Säfken et al., 2021) packages in R 4.4.1 (R Core Team, 2019). All categorical variables were sum-coded the same way as in Experiment 2. A simplified random effects structure with random by-participant and by-item intercepts was used as the model with maximal random effects did not converge. Where interactions were found, data was broken down, and mixed-effects linear models were fitted to subsets of data to explore how the effect of grammaticality was modulated by the other two factors.

4.1.4.2 Eye-tracking data analysis

4.1.4.2.1 Divergence point analysis Participants' accuracy in the comprehension questions were at ceiling level (mean accuracy = 98, $sd = 0.03$), and no participants were excluded from data analysis for low accuracy.

Data were processed in R 4.4.1 (R Core Team, 2019). Using the same principles as Experiment 1, we excluded trials where the participant rated at least one of the classifier-noun pairs as unfamiliar (excluded trials = 163, 8.61% of all trials), as well as trials where no looks on either object were observed for over 50% of the trial duration (excluded trials = 95, 5.02% of all trials). The remaining 1634 trials (86.36% of all trials) were analysed.

Observations where the participant was not fixating on either of the objects were excluded. The remaining data were put into 20ms bins. The same bootstrapping procedure as Experiment 1 was used to determine the onset of fixations to the target object, i.e. the divergence point.

4.1.4.2.2 Applying Bayes' theorem to divergence points Similar to Experiment 1, we applied Bayesian principles to the distribution of divergence points, in order to assess the strength of evidence for or against the $H0$, when results from both Experiment 1 and the current experiment are considered (Stone, 2021). The posterior distributions obtained for Experiment 1 were taken as prior distributions for the current analysis. Again, distributions of divergence point from the divergence point analysis were taken as the likelihood functions, and posterior distributions were obtained using Bayes' theorem.

Again, Bayes factors BF10 were calculated for both T3 sandhi trials and *yi* sandhi trials, using prior distributions of the *difference* in divergence point between conditions (results of Experiment 1), and the posterior distributions of the difference.

4.1.4.3 Correlation between tone sandhi sensitivity and eye movement Since results from Experiment 2 revealed a large variability in the ratings given to T3 sandhi violations in the sandhi-inducing context (e.g., *liang3ben3*), we carried out an additional individual difference analysis to determine whether there is a correlation between the extent to which each participant used the T3 sandhi to generate predictions and their sensitivity to violations of the T3 sandhi.

First, we determined each participant's T3 sandhi sensitivity in the sandhi-inducing context by subtracting their average rating in the ungrammatical condition (T3T3) from the grammatical condition (T2T3). Next, we determined the extent to which each participant used the T3 sandhi to generate predictions in the eye-tracking experiment. For each participant, we took their eye gaze data for the T3 sandhi trials and calculated the difference in their arcsine-transformed looks to the target object between the Different Tones and Same Tones conditions in a time window from 560ms to 960ms after numeral onset (this is equivalent to 200ms before and after the divergence point calculated for the original data before the bootstrapping procedure, averaged across conditions). We took a larger difference between conditions in the looks to the target object to indicate more prediction generated based on the T3 sandhi. Finally, we computed Pearson's correlation coefficient to determine whether there was a significant correlation between participants' T3 sandhi sensitivity and the extent to which participants used the T3 sandhi to predict the target in the eye-tracking experiment.

4.2 Results

4.2.1 Language experience

Similar to Experiment 2, participants' awareness and their reported classroom experience did not differ significantly between the T3 sandhi and the *yi* sandhi in the present experiment (Table 3, classroom experience: $X^2(4) = 4.97$, $p = 0.29$; awareness: $X^2(2) = 2.09$, $p = 0.35$).

	the T3 sandhi			the *yi* sandhi		
	yes	no	do not remember	yes	no	do not remember
explicit knowledge	30 (69.8%)	13 (30.2%)		31 (72%)	12 (28%)	
classroom experience	21 (48.8%)	12 (28%)	10 (23.3%)	20 (46.5%)	7 (16.3%)	16 (37.2%)

Table 3: Responses (percentage of answers) to the tone sandhi experience questionnaire, Experiment 3.

4.2.2 Acceptability rating results

4.2.2.1 Linear mixed-effects models Z-transformed acceptability data is shown in Figure 11. The mixed effect linear model revealed significant main effects of sandhi ($\beta = 0.16$, $t = 20.33$, $p < .001$), context ($\beta = -0.16$, $t = -13.89$, $p < .001$), and grammaticality ($\beta = -0.85$, $t = -108.64$, $p < .001$). All the 2-way and 3-way interactions were also significant.

Two linear mixed-effects models were fitted to data from the T3 sandhi trials and the *yi* sandhi trials respectively. For the T3 sandhi trials, there were main effects of context ($\beta = -0.36$, $t = -24.39$, $p < 0.001$) and grammaticality ($\beta = -0.69$, $t = -54.61$, $p < 0.001$). There was also a significant interaction between context and grammaticality ($\beta = -0.37$, $t = -29.6$, $p < 0.001$). Conditional AIC confirmed that the model including both predictors and their interaction best explained the data ($cAIC = 1086.55$; model without interaction: $cAIC = 1715.91$; model without context: $cAIC = 1770$). Separating data in the sandhi-inducing and the non-sandhi-inducing context suggests significant effects of grammaticality in both the sandhi-inducing context ($\beta = -.63$, $t = -17.67$, $p < 0.001$) and the non-sandhi-inducing context ($\beta = -2.13$, $t = -59.49$, $p < .001$), although such effect was smaller in the sandhi-inducing context.

For the *yi* sandhi trials, there was a significant main effect of context ($\beta = 0.04$, $t = 5.17$, $p < 0.001$), a significant main effect of grammaticality ($\beta = -1.01$, $t = -111.3$, $p < 0.001$), and a significant interaction between the two predictors ($\beta = 0.06$, $t = 6.14$, $p < 0.001$). Model selection using the conditional AIC confirmed that the model including both predictors and their interaction best explained the data ($cAIC = 430.71$; model without interaction: $cAIC = 466.87$; model without context: $cAIC = 491.03$). Further analyses suggest significant effects of grammaticality in both contexts (sandhi inducing: $\beta = -2.13$, $t = -82.95$, $p < 0.001$; non-sandhi-inducing: $\beta = -1.92$, $t = -74.27$, $p < 0.001$), although this effect was slightly larger for the sandhi inducing context.

Overall, results suggest that listeners were highly sensitive to tone sandhi violations in general, as indicated by the significant main effect of grammaticality overall and in all subsets of data. Phonological context played a role in listeners' acceptability ratings, as well. While for the *yi* sandhi trials, listeners found its violation slightly less unacceptable in the non-sandhi-inducing context; for the T3 sandhi trials, listeners found tone sandhi violation much more acceptable in the sandhi-inducing context, namely, phrases such as *liang3ben3* were more acceptable than *liang2ge4*. However, despite phonological context having an influence on the effect size of grammaticality, listeners still rated ungrammatical trials significantly lower than grammatical trials, for both the *yi* sandhi and the T3 sandhi, in both the sandhi-inducing and the non-sandhi-inducing contexts.

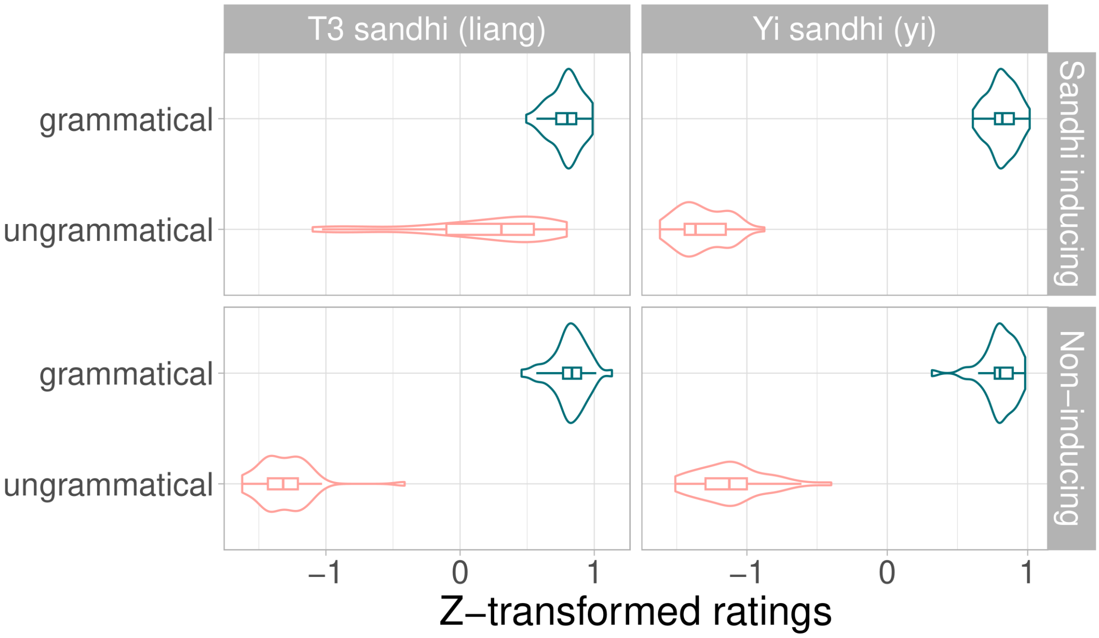

Figure 11: Z-transformed acceptability ratings by condition, aggregated by subject, Experiment 3. Scattered points show by-subject mean score.

4.2.3 Eye-tracking results

4.2.3.1 Divergence point analysis The proportion of looks to the target and competitor objects is shown in Figure 12. Figure 13 shows the distribution of divergence point differences.

For trials with the T3 sandhi, the estimated divergence point was 819ms in the Different Tones condition with a 95% confidence interval of [780, 860], and 765ms [720, 820] in the Same Tones condition. The divergence point was 54ms later in the Different Tones condition compared to the Same Tones condition, with the 95% CI of [-20, 120] covering both positive and negative values, suggests no significant difference between conditions in terms of the onset of looks to the target object.

For trials with the *yi* sandhi, the estimated divergence point 686ms [620, 740] in the Different Tones condition, and 654ms [620, 720] in the Same Tones condition. The divergence point was 32ms later in the Different Tones condition compared to the Same Tones condition, with a 95% CI of [-60, 100] covering both negative and positive values, suggesting no effect of prediction.

Overall, results from the divergence point analysis showed that listeners were not any faster to look at the target object in the Different Tones condition than in the Same Tones condition, for both the T3 sandhi trials and the *yi* sandhi trials.

4.2.3.2 Applying Bayes' theorem to divergence points Posterior distributions of divergence points were $\mathcal{N}(781, 18)$ for T3 sandhi trials in the Different Tones condition and $\mathcal{N}(753, 18)$ in the Same Tones condition. They were $\mathcal{N}(641, 20)$ for *yi* sandhi trials in the Different Tones condition and $\mathcal{N}(635, 18)$ in the Same Tones condition. The posterior *difference* between conditions were $\mathcal{N}(28, 25)$ for T3 sandhi trials and $\mathcal{N}(6, 27)$ for *yi* sandhi trials (Figure 14).

Bayes factors calculated using the Savage-Dickey equation quantify the extent to which the data shift belief toward the null hypothesis ($H0$) versus the alternative ($H1$). For the T3 sandhi trials, a Bayes factor $BF10 = 0.14$ indicates a moderate shift in favour of the $H0$, suggesting that listeners did not use T3 sandhi cues to generate lexical predictions. Meanwhile, for the *yi* sandhi trials, a Bayes factor $BF10 = 0.73$ indicates a weak shift toward $H0$. This weakness of the shift is expected, given that the prior distribution for the difference in divergence points was already concentrated near zero.

Overall, Bayes factors suggest an increased degree of support for the null hypothesis across both the T3 sandhi and *yi* sandhi trials.

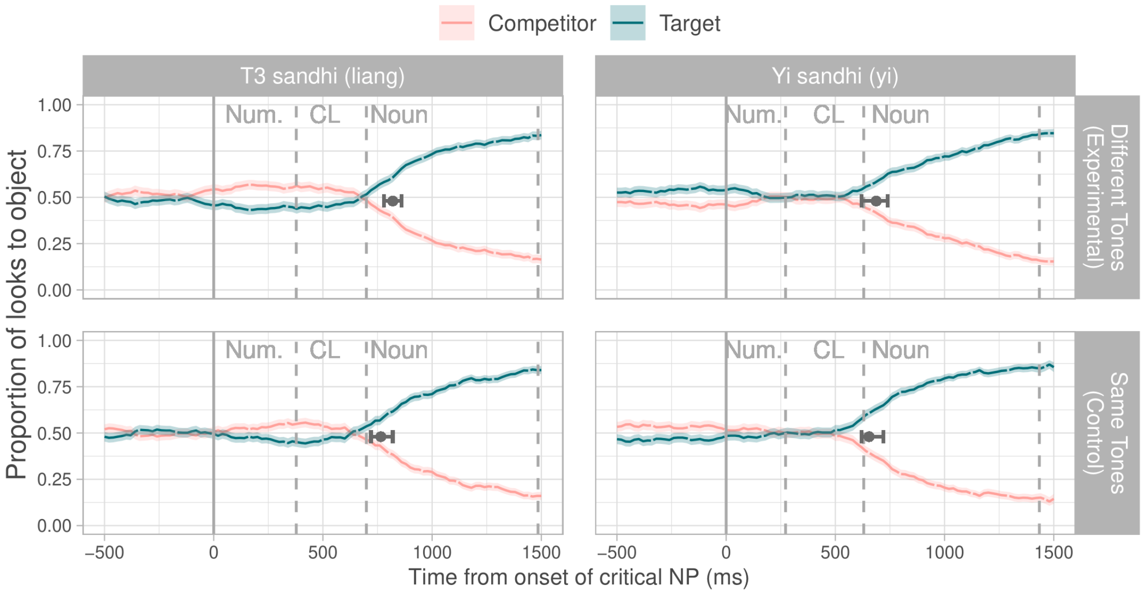

Figure 12: Proportion of looks to the target and competitor objects, Experiment 3 (Left column: T3 Sandhi; Right column: Yi Sandhi; Top row: Different Tones condition; Bottom row: Same Tones condition). The solid vertical line shows the onset of the critical NP (i.e. the onset of the numeral). The dashed lines show the average onset of the classifier and noun, as well as the offset of the noun. The solid point represents the mean divergence point estimated from the bootstrapping procedure, the error bars represent the 95% confidence interval.

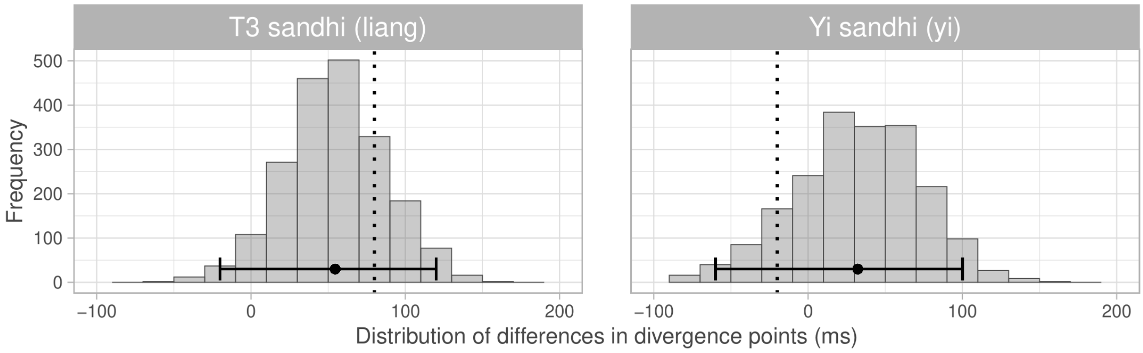

Figure 13: Bootstrapped distribution of differences in divergence points between the conditions (Different Tones minus Same Tones) in the T3 sandhi (left) and *yi* sandhi trials (right), Experiment 3. The solid point and the error bars represent the mean difference and the 95% confidence interval respectively. The vertical dotted lines represent the difference calculated from doing the estimation procedure on the original data before resampling.

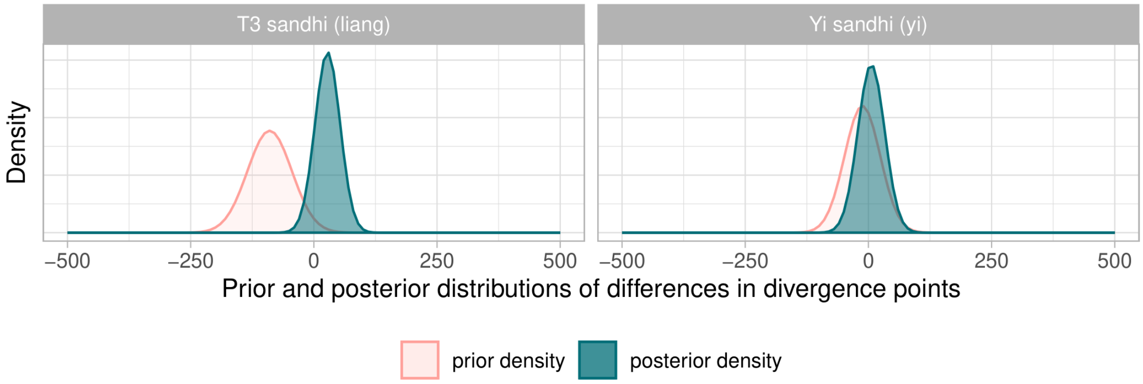

Figure 14: Prior and posterior distributions of differences in divergence points in the T3 sandhi (left) and *yi* sandhi trials (right), Experiment 3.

4.2.4 Correlation between tone sandhi sensitivity and eye movement

Figure 15 shows the by-participant mean ratings to the T3 sandhi Ungrammatical trials in the sandhi-inducing context (T3T3 trials) in the present experiment. The individual difference analysis revealed no significant correlation between participants' T3 sandhi sensitivity and their arcsine-transformed difference in looks to the target in T3 sandhi trials ($r = 0.09$, $t(40) = 0.43$, $p = 0.56$, Figure 16). The non-significant correlation indicates that sensitivity to T3 sandhi violations did not modulate the extent to which the listeners use the T3 sandhi to generate predictions.

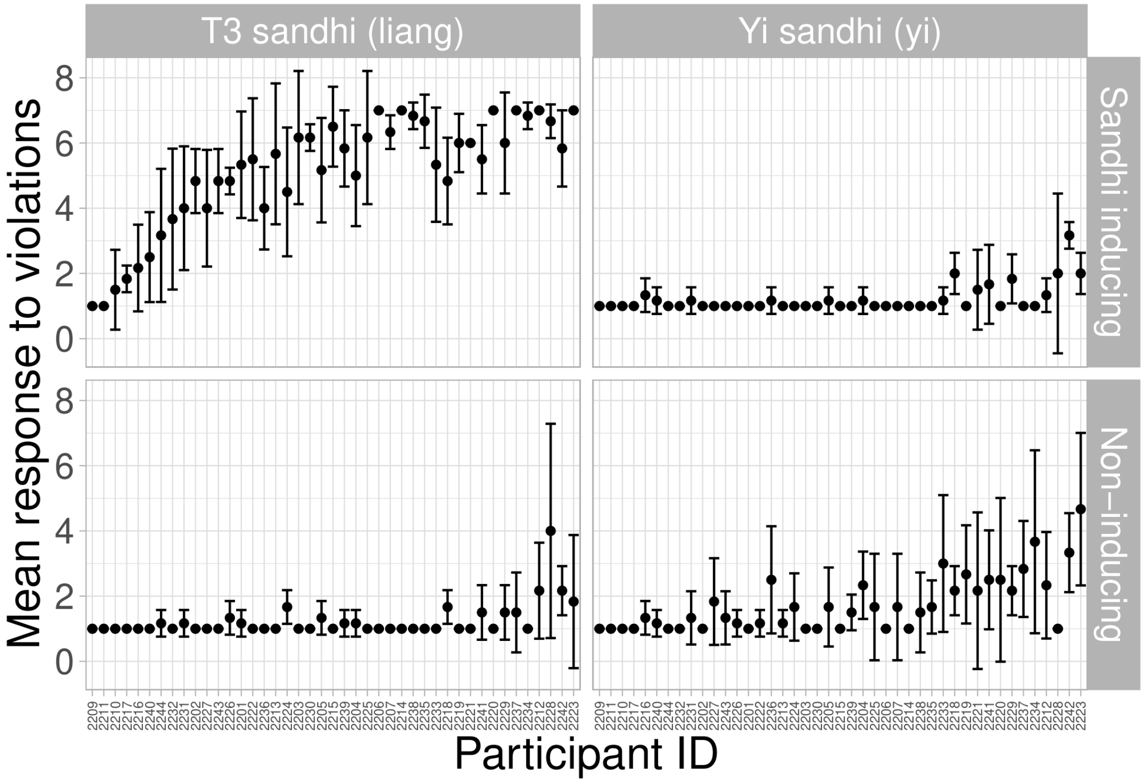

Figure 15: By-subject mean acceptability ratings given to ungrammatical trials, Experiment 3. Ordered by average ratings given to ungrammatical trials. Error bars show standard deviations.

4.2.5 Separating *liang2* and *liang3* trials

Following Experiment 1, we separately analysed experimental items where the base tone *liang3* was heard in the Different Tones condition vs. where the sandhi tong *liang2* was heard in the Different Tones condition (Figure 17).

For the items with the base tone *liang3* in both conditions, divergence point analysis revealed a mean onset of looks to the target object of 783ms [740, 840] in the Different Tones condition, and 812ms [740, 900] in the Same Tones condition. There is a mean difference of -29ms [-140, 60] between conditions (Different Tones - Same Tones), suggesting no evidence that the onset of looks to the target object was earlier in the Different Tones condition than in the Same Tones condition.

For the items with the sandhi tone *liang2* pronounced in the Different Tones condition and the base tone *liang3* in the Same Tones condition, the mean onset of looks to the target was 920ms [860, 1020] in the Different Tones condition and 815ms [760, 880] in the Same Tones condition. A between-condition difference of 405ms [0, 220] suggests no evidence for an earlier onset in the Different Tones condition (Figure 18).

Overall, these results suggest that listeners could not use the T3 sandhi in numerals to generate predictions of the upcoming classifier and noun, regardless of whether they heard the base or sandhi tone.

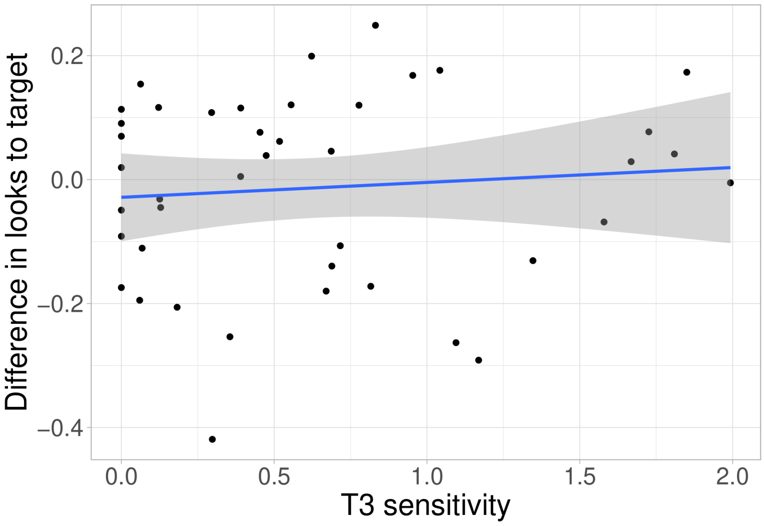

Figure 16: Correlation between participants' sensitivity to T3 sandhi violations (ratings for grammatical trials - ratings for ungrammatical trials) and the difference between their arcsine-transformed proportion of looks in the Different Tones and the Same Tones conditions in the critical window of T3 sandhi trials. Scattered points represent by-participant means.

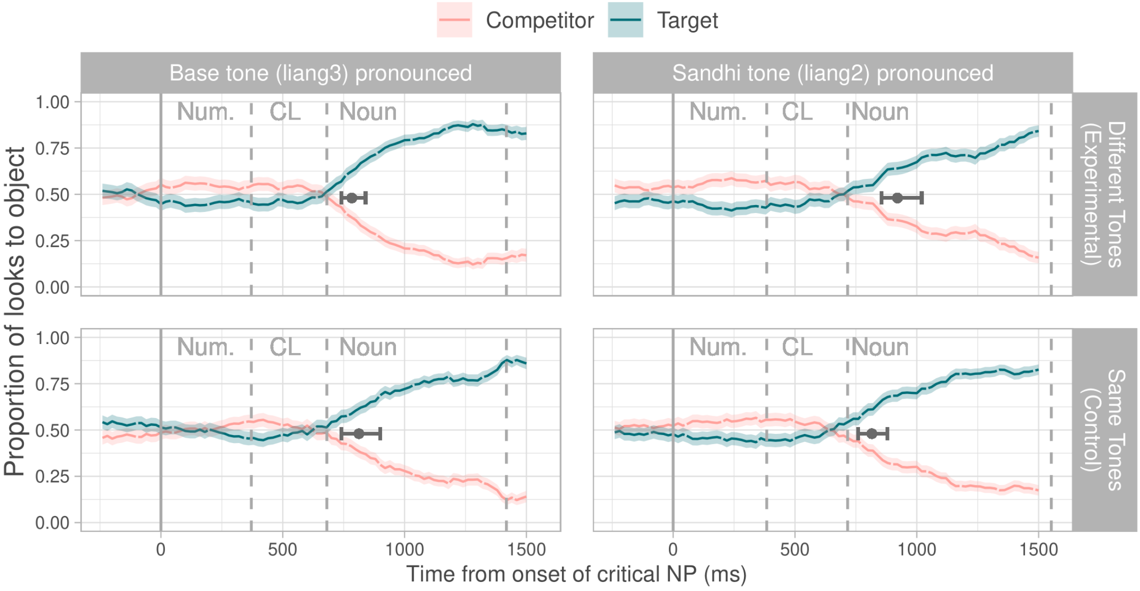

Figure 17: Proportion of looks to the target and competitor objects, Experiment 3, T3 sandhi trials. Left column: trials with *liang3* in both conditions; Right column: trials with *liang2* in the Different Tones condition and *liang3* in the Same Tones condition; Top row: Different Tones condition; Bottom row: Same Tones condition.

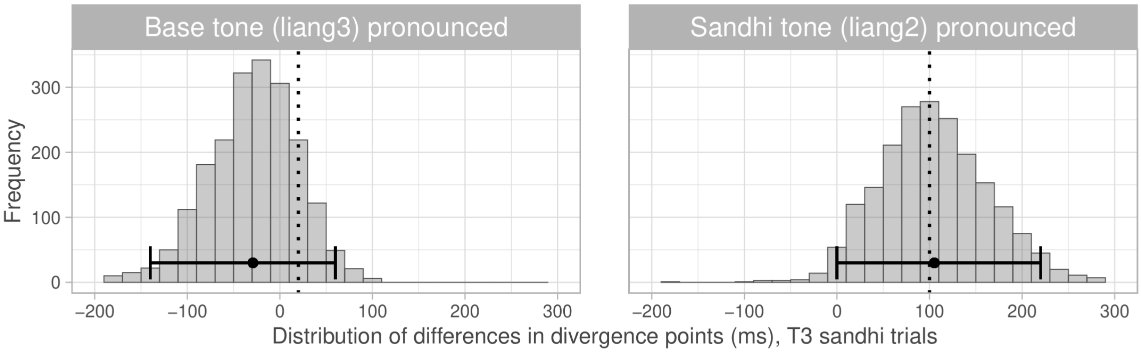

Figure 18: Bootstrapped distribution of differences in divergence points between the conditions (Different Tones minus Same Tones), Experiment 3, T3 sandhi trials. Left: trials with *liang3* in both conditions; Right: trials with *liang2* in the Different Tones condition and *liang3* in the Same Tones condition.

4.3 Discussion

The current experiment was a direct replication of Experiments 1 and 2, collecting data from a single group of participants. We tested whether listeners of Mandarin Chinese could use the *yi* sandhi and the T3 sandhi in numerals to predict an upcoming classifier and noun, and whether there is a correlation between the extent to which a listener uses the T3 sandhi in prediction and their sensitivity to T3 sandhi violations.

As in Experiment 2, listeners showed clear sensitivity to violations of both tone sandhi patterns in the acceptability rating task. Further, there was a large variability in listeners' sensitivity to T3 sandhi violations in the sandhi-inducing context.

In the eye-tracking experiment, meanwhile, we found that listeners did not look toward the target object earlier in the Different Tones condition compared to the Same Tones condition in either T3 or *yi* sandhi trials. This contrasts with the findings from Experiment 1, where listeners demonstrated predictive effects in T3 sandhi trials, but not in *yi* sandhi trials. Using the findings from Experiment 1 as priors, we analysed the results of the current experiment following Bayesian principles and found support for the null hypothesis (H0), suggesting no significant difference in the onset of looks to the target object across conditions for both tone sandhi patterns. This implies that, when combining results from Experiment 1 and the current experiment, listeners overall could not use either the T3 sandhi or *yi* sandhi pattern to predict the upcoming classifier and noun.

Further, following findings from Experiment 1, we conducted an additional analysis focusing on the trials where the numeral *liang* was realised in its sandhi form *liang2* in the Different Tones condition. Regardless of the tone that the numeral was realised in, we did not find evidence for their impact on listeners' prediction, as listeners' eye-movements were not modulated by the tone sandhi cue in the numeral. Bayes Factors also supported the absence of a prediction effect.

In the current experiment, we also explored the relationship between listeners' sensitivity to tone sandhi violations and their ability to generate predictions using tone sandhi. As the variability in sensitivity was the most prominent for T3T3 sequences, we tested whether there is a correlation between individuals' sensitivity to T3T3 sequences and the extent to which they used the T3 sandhi to predict. We did not find a significant correlation between the two, suggesting that listeners' inability to use tone sandhi cues to predict is not attributable to a lack of sensitivity to the tone sandhi patterns.

Taken together, the results of the current study suggest that even though listeners of Mandarin Chinese have robust knowledge of both tone sandhi patterns, they cannot use the tone sandhi cues in numerals to predict the upcoming classifiers and nouns on the fly.

5 General discussion

In this study, we used Mandarin Chinese tone sandhi as a test case to investigate whether listeners use phonological knowledge as an input to prediction during language comprehension. In Experiment 1, we showed initial evidence from a visual world eye-tracking paradigm that listeners of Mandarin Chinese may be able to use the T3 sandhi, but not the *yi* sandhi, in a numeral to predict an upcoming classifier and noun. In Experiment 2, we demonstrated that listeners are perceptually highly sensitive to both tone sandhi patterns and their violations in an acceptability judgment task, indicating that they have robust knowledge about both tone sandhi patterns.

This suggests that listeners' inability to use the *yi* sandhi to predict in Experiment 1 could not be attributed to their insensitivity to said sandhi. Instead, we found that participants were on average less sensitive to T3 sandhi violations in a sandhi-inducing context (T3T3 sequences) and there was much variability in participants' ratings of such violations. In Experiment 3, we collected eye movement data and acceptability data from a single group of participants and investigated whether there was a correlation between listeners' sensitivity to the T3 sandhi and their ability to use it to compute predictions. Eye-tracking results suggested that listeners did not use either sandhi pattern to predict the upcoming classifier and noun. Applying Bayesian principles to divergence points also revealed no effects of prediction for either sandhi patterns when data from Experiment 1 and Experiment 3 were considered together. Further, listeners' T3 sandhi sensitivity did not correlate with the extent to which they can use it in prediction. In summary, the current study suggests that although native listeners of Mandarin Chinese have clear knowledge of both tone sandhi patterns, they do not seem to use it in computing predictions about upcoming language on the fly.

5.1 Tone sandhi cues as a (poor) input to lexical prediction

At first glance, these results seem incompatible with recent findings by Liu et al. (2025). In Liu et al. (2025)'s study, listeners showed anticipatory eye movements toward target objects when the target phrase was preceded by an informative *yi* sandhi cue. The design of the current study (Experiments 1 and 3) shares several key features with that of Liu et al. (2025). Both studies employed the visual world eye-tracking paradigm, manipulated the presence of an informative tone sandhi cue across conditions, and asked whether listeners looked towards the target object earlier when said cue was present.

However, beyond these similarities, the current study differed from Liu et al. (2025)'s study in a few important ways. Liu et al. (2025)'s experiment only investigated the *yi* sandhi and the numeral *yi* was present in all trials. In addition, the same sentence frame ("On the screen, is there *yi*...?") was used repeatedly throughout the experiment. Both these properties likely made the presence of the informative cue *yi* highly salient and predictable. In contrast, the current study used a variety of sentence frames to investigate two tone sandhi patterns in two different numerals (*yi* and *liang*). In addition, the current study included a set of filler items that contained no numerals or numerals other than *yi* or *liang.* The variability in the current study's stimuli ensured that participants could not predict the presence or the position of an informative numeral.

Additionally, the study by Liu et al. (2025) used an artificially slow stimulus presentation rate (800ms per syllable) which included noticeable silent pauses between syllables. This could also have heightened the salience of *yi* as an informative cue to the identity of the target. The current study, in contrast, used a naturalistic speech rate without noticeable silence between syllables. The naturalistic speech rate ensured the ecological validity of the results as the spoken language input was similar to what participants would encounter in real life.

In sum, the salience of the informative tone sandhi cue differed notably between the current study and Liu et al. (2025). While Liu et al. (2025)'s experiment took a variety of measures to artificially increase the saliency of *yi* sandhi cues, the current study took measures to minimise that to study whether and how listeners may use tone sandhi cues for prediction during real-time language comprehension. Given *yi* sandhi cues' high salience in Liu et al. (2025)'s study, it is plausible that participants in their study were more attuned to the occurrence of *yi*, allowing them to develop experiment-specific strategies to use its tone for lexical predictions. In the current study, however, the less predictable sentential context and naturalistic stimuli presentation likely reduced the likelihood of developing such strategies, which may explain the contrasting results.

Meanwhile, the observed null prediction effects also seem to somewhat differ from the results in A. Ito & Hirose (2024). In that study, listeners were able to select a target object more quickly if the noun was preceded by an informative Kansai Japanese (KJ) pitch accent sandhi cue. The authors proposed that this facilitation effect may be compatible with the hypothesis that KJ listeners used pitch accent sandhi cues to restrict the set of possible upcoming tones.

However, as the authors later pointed out, it is worth noting that although A. Ito & Hirose (2024) found a reaction time (RT) facilitation in the presence of informative pitch accent sandhi cues, the RTs never preceded target onset. This suggests that the observed effect may not necessarily indicate prediction. Further, this facilitation effect was not limited to native Kansai speakers—it was also observed in the control group of Standard Japanese speakers who had minimal exposure to the Kansai dialect and no knowledge of Kansai pitch accent sandhi, serving as a control group. The authors attributed the effects in the control group to a universal constraint against a sequence of low tones. Given that both groups showed similar facilitation, Occam's razor suggests a parsimonious explanation: the same mechanism, a universal constraint against a sequence of low tones as discussed by the authors, likely underlies the results for both groups.

In sum, neither Liu et al. (2025) nor A. Ito & Hirose (2024) provided unequivocal evidence that phonological cues regularly feed into lexical predictions during language comprehension. The present results build on this by offering clearer evidence that listeners do not use tone sandhi cues to predict upcoming words, suggesting that such predictions may be either infrequent or constrained by specific conditions.

5.2 Challenges in detecting prediction in the visual world paradigm

Next, we discuss potential explanations for listeners' failure to use tone sandhi cues to make predictions about the upcoming lexical items in the present study. In the present study, the tone of the numeral was only directly predictive about the tone of the syllable that followed (the classifier), and not the noun (or the tone of the noun) itself. In other words, listeners must be able to connect the tone of the numeral to the tone of the classifier, and the classifier to the noun in order to identify the target object. The indirect connection between what can be predicted (the tone of the classifier) and the objects in the visual display may have introduced challenges in detecting prediction effects through eye movements in the current study.

Specifically, several conditions must be met in order for us to detect prediction effects in the present study. First, the visual display needed to activate the labels associated with the depicted objects (the nouns, e.g. *chuang2* for the display of a bed and *deng4zi* for the display of a stool). Next, participants needed to activate the classifiers that are associated with the two nouns and their phonological form including their tone (e.g. that *chuang2* "bed" takes on the classifier *zhang1*, and *deng4zi* "stool" takes on *ba3* or occasionally the general classifier *ge4*). Then, upon hearing the numeral in the input, listeners need to recognise it as a numeral, identify whether it has undergone tone sandhi, apply their knowledge of tone sandhi to determine what tone(s) the following syllable can be in, and use this information to select the suitable classifier (and the noun associated with it). A failure to complete any of these steps would prevent the listeners from showing a predictive effect.

For instance, listeners may associate the labels of the depicted objects with different classifiers from the ones intended by the experiment (e.g. the noun *ji2ta1* (guitar) is associated with the classifier *ba3*, but it can also be used with the general classifier *ge4*). If listeners activated only the classifier *ba3* with the noun *ji2ta1*, then hearing a T3 numeral (e.g. *liang3*) would rule out *ji2ta1* as the target. If, however, the listeners activated both *ba3* and *ge4* for the noun *ji2ta1*, then numerals such as *liang3* would not be informative because it would still be compatible with the classifier *ge4*.

We have taken two key measures to minimise the impact of this. First, we only used pictures for which more than 60% of participants in the norming study agreed on the label (noun) as well as the classifier. Second, we excluded from the data analysis all trials in which the participant would not use the classifier in the stimuli with the given noun.

Another possibility is that listeners may fail to activate any classifiers upon seeing the objects in the visual display. In this case, even if listeners can use the tone of the numeral to predict an upcoming lexical tone (e.g. *liang2* suggests an upcoming T3 word), they may not be able to identify the target if the classifiers were not already activated by the visual display. To address the issue of classifier activation more directly, future studies could present the classifiers along with the nouns explicitly—such as by presenting the whole noun phrase in written form on the visual display—thereby eliminating the extra step that listeners need to make to connect what are on the visual display (the objects) and what they can predict using the tone sandhi cues (the tone of the classifier).

To further avoid the classifier's impact, future studies can investigate tone sandhi outside numerals and classifiers, for instance the T3 sandhi in verb phrases such as *da3-che1* hit-car "call a taxi" vs. *da2-gu3* hit-drum "play the drum" or noun phrases such as *xiao3-mao1* small-cat "small cat/kitten" vs. *xiao2-gou3* small-dog "small dog/puppy".

In sum, although the indirect connection between what can be predicted (the tone of the classifier) and the objects on the visual display may have made it more difficult to observe effects compared to the typical visual world paradigm, we have taken different measures to maximise the observability of potential prediction effects. Therefore, the results of the current study still offer valuable insights into the minimal role tone sandhi cues play in generating lexical predictions.

5.3 Why Mandarin tone sandhi may differ from other phonological cues

The results of the current study do not suggest that prediction based on phonological cues never occurs. For instance, K. Ito & Speer (2008) found that informative English contrastive pitch accents can lead to anticipatory eye movements toward target objects in an object selection task (e.g. "*Pick up the blue ball. Now, pick up the*

GREEN ball"). Similarly, Weber et al. (2006) showed that contrastive accents in adjectives (e.g., "*Please hand me the RED scissors/vase*") led to anticipatory fixations on contrastive targets (e.g., "*RED scissors*" when multiple scissors were present) but not on noncontrastive ones (e.g., "*RED vase*" when only one vase was present). In both cases, listeners used contrastive pitch accents to anticipate repeated nouns or to infer the necessity of specification.

English contrastive pitch accent carries important cues about information structure and indicates to the listener what is new information in the sentence and should be focused on. Consequently, contrastive pitch accents are directly informative about the identity of the upcoming lexical item rather than merely its phonological form. For example, upon hearing an informative contrastive pitch accent ("*GREEN*"), the listener can infer that the upcoming referent is likely something that has previously been named ("*ball*"). This relatively direct mapping between the predictive cue and the target at the lexical level makes lexical prediction based on contrastive pitch accents relatively straightforward.

In contrast, Mandarin Chinese tone sandhi cues, as examined in the current study, is not informative about upcoming language input on the lexical level. Instead, these cues are only informative about the tonal form of the upcoming input. For example, *liang2* indicates that the following syllable will be in T3 (and, given its position as a numeral, likely a T3 classifier). However, there are many T3 classifiers, each lexically distinct and carrying different semantic features. Thus, although tone sandhi cues restrict the set of possible upcoming words (by excluding classifiers in other tones), the restrictions they impose is not semantic or lexical in nature. Therefore, tone sandhi information may be an ineffective cue to predict upcoming words in naturalistic settings.

In a visual world paradigm, the situation is somewhat different from naturalistic language use. Crucially, the set of possible upcoming lexical items is constrained by the visual display. In the present study, for example, there were only two objects displayed on the screen, and one of them was always referred to in the spoken sentence. This significantly reduces the computational demands required to generate lexical predictions using tone sandhi cues, compared to in naturalistic contexts where the range of possible upcoming words is arguably much wider. Despite this simplification, the results of the current study suggest that lexical prediction based on tone sandhi remains difficult to observe.

5.4 Linguistic knowledge vs. real-time lexical predictions

An interesting contrast in this study lies between listeners' sensitivity to sandhi violations and their seeming inability to use tone sandhi knowledge to make lexical predictions during real-time comprehension. In the acceptability judgment task (in Experiments 2 and 3), we found that native speakers are highly sensitive to violations of both tone sandhi rules, indicating robust knowledge of the underlying phonological patterns. However, in the eye-tracking experiments (Experiments 1 and 3), we struggled to find evidence that listeners utilize this knowledge to predict upcoming words during comprehension.

One possible explanation for this is that although listeners have robust linguistic knowledge of tone sandhi, their predictive mechanism is "deaf" to these patterns during real-time language comprehension. This proposal is somewhat out of line with the broader literature on real-time sentence processing. Studies have shown that comprehenders are sensitive to a wide range of grammatical constraints in real-time comprehension. For example, when processing sentences with long-distance dependencies, adult comprehenders are shown to expect the occurrence of a gap at upcoming grammatical locations once they encounter a filler, i.e. predicting an upcoming syntactic structure (e.g. Lee, 2004; Stowe, 1986; Traxler & Pickering, 1996). Similarly, comprehenders use a wide range of linguistic information including sentential and discourse context (Altmann & Kamide, 1999; Federmeier & Kutas, 1999; Kutas & Hillyard, 1984; Otten & Van Berkum, 2008; Van Berkum et al., 2005) as well as linguistic markers of gender (Lew-Williams & Fernald, 2007, 2010; Stone, Verı́ssimo, et al., 2021) and case (Kamide, Scheepers, et al., 2003) to generate relevant expectations about upcoming language input. The general pattern here is that listeners tend to use a wide array, if not all, of their grammatical knowledge to anticipate upcoming language input. Thus, our findings about tone sandhi constitute a clear exception to this pattern.

In terms of phonological processing, previous research has also shown that listeners can anticipate upcoming phonemes based on informative phonological cues such as English nasal assimilation (Gaskell & Marslen-Wilson, 1996; Gow Jr, 2001; Lahiri & Marslen-Wilson, 1991; Otake et al., 1996). Similarly, there is a possibility that comprehenders can compute a *tonal* prediction based on a tone sandhi cue (e.g. expecting an upcoming T3 when hearing a sandhi tone T2), but they may not be able to map this tonal expectation onto specific lexical items. Future studies could test whether listeners establishing any expectations on the phonological level based on tone sandhi cues using tasks such as lexical tone monitoring.

5.5 From tonal form to lexical identity

Similar to what is found about informative nasal assimilation cues, it is possible that Mandarin Chinese listeners are able to use informative tone sandhi cues to anticipate an upcoming tone. However, having expectations about the upcoming syllable's tone may not translate into lexical predictions. Listeners must perform additional computations to generate a lexical prediction based on their expectation about the upcoming syllable's tone. In the visual world paradigm, this involves matching the expected tone with possible sentence continuations indicated by the visual display and identifying which continuations are possible.

This additional mapping step may be a key reason why lexical prediction based on tone sandhi is rare or difficult. Such additional computations from a phonological expectation to a lexical prediction may be time- and resource-consuming. In real-life language use, where contextual information does not strongly restrict the set of possible upcoming words, converting expectations about phonological form to lexical predictions would be even more challenging, as many lexical items with varying semantic features can match the expected phonological form. As a result, while listeners may routinely generate phonological-level predictions from informative phonological cues, these predictions may rarely project to the lexical level. This may carry over to the visual world eye-tracking paradigm, where although the visual display restricts the target to a few possible candidates (two in the case of the current study), listeners may not be able to immediately adapt to this constraining non-linguistic context and the projection from a phonological expectation to a lexical prediction may still not readily occur.

5.6 Prediction and predictability

If lexical predictions based on tone sandhi cues are rare, a related question is whether informative tone sandhi cues influence lexical predictability. It seems intuitive that words compatible with an informative cue should become more likely or more predictable, while those that are incompatible should become less likely or activated. However, if comprehenders rarely or never engage in lexical prediction based on tone sandhi cues, it is plausible that hearing an informative tone sandhi cue has little to no effect on the predictability of the upcoming word.

Determining the effect of informative tone sandhi cues on lexical predictability could also help reconcile the divergent findings between the current study and Liu et al. (2025). If tone sandhi cues significantly affect lexical predictability, it will suggest a gap between real-time and offline language processing where certain mechanisms may not be active under the time-pressure of naturalistic speech. Conversely, if tone sandhi cues have minimal impact on predictability, it is highly likely that participants in Liu et al. (2025)'s study employed experiment-specific strategies that are unlikely to occur in naturalistic settings.

In psycholinguistics, predictability is often measured using cloze probability in sentence completion tasks, where participants provide the most likely continuation of a sentence. To our best knowledge, no study has examined whether tone sandhi cues affect cloze probability. Future research could address this gap by investigating whether tone sandhi influences comprehenders' responses in an offline sentence completion task. This approach would help determine whether tone sandhi naturally feeds into lexical predictions, offering valuable insights into how phonological information shapes predictive processes in language comprehension.

6 Conclusions

The present study used Mandarin Chinese tone sandhi patterns in numerals to investigate whether listeners make use of phonological information in a numeral to anticipate an upcoming classifier and noun. We tested two tone sandhi patterns: T3 sandhi with the numeral*liang* meaning "two", and *yi* sandhi with the numeral *yi* meaning "one". Although results from an acceptability task showed that listeners were perceptually highly sensitive to both tone sandhi patterns; the overall evidence from two visual world eye-tracking experiments suggested that listeners were not able to use tone sandhi cues to predict upcoming words on the fly. Our results indicate that listeners may not regularly use tone sandhi cues to generate lexical predictions.

References

Altmann, G. T., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*(3), 247–264.

Altmann, G. T., & Kamide, Y. (2007). The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language*, *57*(4), 502–518.

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, *52*, 388–407.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412.

Bar, M. (2011). *Predictions in the brain: Using our past to generate a future*. Oxford University Press.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278.

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

Boersma, P., & Weenink, D. (2022). *Praat: Doing phonetics by computer* (6.2.14). http://www.praat.org/

Cai, Q., & Brysbaert, M. (2010). SUBTLEX-CH: Chinese word and character frequencies based on film subtitles. *PLOS ONE*, *5*(6), 1–8. https://doi.org/10.1371/journal.pone.0010729

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

Chen, A., Liu, L., & Kager, R. (2015). Cross-linguistic perception of mandarin tone sandhi. *Language Sciences*, *48*, 62–69.

Chen, J., Bowerman, M., Huettig, F., & Majid, A. (2010). Do language-specific categories shape conceptual processing? Mandarin classifier distinctions influence eye gaze behavior, but only during linguistic processing. *Journal of Cognition and Culture*, *10*(1-2), 39–58.

Chow, W.-Y., & Chen, D. (2020). Predicting (in) correctly: Listeners rapidly use unexpected information to revise their predictions. *Language, Cognition and Neuroscience*, *35*(9), 1149–1161.

Chow, W.-Y., Smith, C., Lau, E., & Phillips, C. (2016). A "bag-of-arguments" mechanism for initial verb predictions. *Language, Cognition and Neuroscience*, *31*(5), 577–596.

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, *8*(8), 1117–1121.

Duñabeitia, J. A., Crepaldi, D., Meyer, A. S., New, B., Pliatsikas, C., Smolka, E., & Brysbaert, M. (2018). MultiPic: A standardized set of 750 drawings with norms for six european languages. *Quarterly Journal of Experimental Psychology*, *71*(4), 808–816.

Federmeier, K. D., & Kutas, M. (1999). A rose by any other name: Long-term memory structure and sentence processing. *Journal of Memory and Language*, *41*(4), 469–495.

Gaskell, M. G., & Marslen-Wilson, W. D. (1996). Phonological variation and inference in lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, *22*(1), 144.

Gow, David W., & McMurray, B. (2007). Word recognition and phonology: The case of english coronal place assimilation. In J. Cole & J.-I. Hualde (Eds.), *Papers in laboratory phonology* (Vol. 9, pp. 173–200). Mouton de Gruyter.

Gow Jr, D. W. (2001). Assimilation and anticipation in continuous spoken word recognition. *Journal of Memory and Language*, *45*(1), 133–159.

Grüter, T., Lau, E., & Ling, W. (2018). L2 listeners rely on the semantics of classifiers to predict. *Proceedings of the 42nd Annual Boston University Conference on Language Development*, 303–316.

Hallé, P. A., Chang, Y.-C., & Best, C. T. (2004). Identification and discrimination of mandarin chinese tones by mandarin chinese vs. French listeners. *Journal of Phonetics*, *32*(3), 395–421.

Huettig, F., Chen, J., Bowerman, M., & Majid, A. (2010). Do language-specific categories shape conceptual processing? Mandarin classifier distinctions influence eye gaze behavior, but only during linguistic processing. *Journal of Cognition and Culture*, *10*(1-2), 39–58.

Ito, A. (2024). Phonological prediction during comprehension: A review and meta-analysis of visual-world eye-tracking studies. *Journal of Memory and Language*, *139*, 104553.

Ito, A., & Hirose, Y. (2024). Sandhi-based predictability of pitch accent facilitates word recognition in kansai japanese speakers. *Quarterly Journal of Experimental Psychology*, *0*(0), 17470218241237219. https://doi.org/10.1177/17470218241237219

Ito, A., Pickering, M. J., & Corley, M. (2018). Investigating the time-course of phonological prediction in native and non-native speakers of english: A visual world eye-tracking study. *Journal of Memory and Language*, *98*, 1–11.

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.

Kamide, Y., Altmann, G. T., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, *49*(1), 133–156.

Kamide, Y., Scheepers, C., & Altmann, G. T. (2003). Integration of syntactic and semantic information in predictive processing: Cross-linguistic evidence from german and english. *Journal of Psycholinguistic Research*, *32*, 37–55.

Klein, N. M., Carlson, G. N., Li, R., Jaeger, T. F., & Tanenhaus, M. K. (2012). Classifying and massifying incrementally in chinese language comprehension. *Count and Mass Across Languages*, 261–282.

Kukona, A. (2020). Lexical constraints on the prediction of form: Insights from the visual world paradigm.

Journal of Experimental Psychology: Learning, Memory, and Cognition, *46*(11), 2153.

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

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, *307*(5947), 161–163.

Kwon, N., Sturt, P., & Liu, P. (2017). Predicting semantic features in chinese: Evidence from ERPs. *Cognition*, *166*, 433–446.

Lahiri, A., & Marslen-Wilson, W. (1991). The mental representation of lexical form: A phonological approach to the recognition lexicon. *Cognition*, *38*(3), 245–294.

Lee, M.-W. (2004). Another look at the role of empty categories in sentence processing (and grammar). *Journal of Psycholinguistic Research*, *33*, 51–73.

Lenth, R. V. (2024). *Emmeans: Estimated marginal means, aka least-squares means*. https://CRAN.R-project.org/package=emmeans

Lew-Williams, C., & Fernald, A. (2007). Young children learning spanish make rapid use of grammatical gender in spoken word recognition. *Psychological Science*, *18*(3), 193–198.

Lew-Williams, C., & Fernald, A. (2010). Real-time processing of gender-marked articles by native and non-native spanish speakers. *Journal of Memory and Language*, *63*(4), 447–464.

Li, C. N., & Thompson, S. A. (1989). *Mandarin chinese: A functional reference grammar*. Univ of California Press.

Li, X., & Chen, Y. (2015). Representation and processing of lexical tone and tonal variants: Evidence from the mismatch negativity. *PloS One*, *10*(12), e0143097.

Li, X., Li, X., & Qu, Q. (2022). Predicting phonology in language comprehension: Evidence from the visual world eye-tracking task in mandarin chinese. *Journal of Experimental Psychology: Human Perception and Performance*, *48*(5), 531.

Liu, S., Chen, X., & Wang, S. (2025). The role of tonal information in speech prediction: Evidence based on chinese tone sandhi. *Language, Cognition and Neuroscience*, 1–17.

Lukyanenko, C., & Fisher, C. (2016). Where are the cookies? Two-and three-year-olds use number-marked verbs to anticipate upcoming nouns. *Cognition*, *146*, 349–370.

Martin, C. D., Thierry, G., Kuipers, J.-R., Boutonnet, B., Foucart, A., & Costa, A. (2013). Bilinguals reading in their second language do not predict upcoming words as native readers do. *Journal of Memory and Language*, *69*(4), 574–588.

Nieuwland, M. S., Politzer-Ahles, S., Heyselaar, E., Segaert, K., Darley, E., Kazanina, N., Von Grebmer Zu Wolfsthurn, S., Bartolozzi, F., Kogan, V., Ito, A., et al. (2018). Large-scale replication study reveals a limit on probabilistic prediction in language comprehension. *ELife*, *7*, e33468.

Otake, T., Yoneyama, K., Cutler, A., & Van Der Lugt, A. (1996). The representation of japanese moraic nasals. *The Journal of the Acoustical Society of America*, *100*(6), 3831–3842.

Otten, M., & Van Berkum, J. J. (2008). Discourse-based word anticipation during language processing: Prediction or priming? *Discourse Processes*, *45*(6), 464–496.

Peng, S.-H. (2000). Lexical versus 'phonological'representations of mandarin sandhi tones. *Papers in Laboratory Phonology V: Acquisition and the Lexicon*, *5*, 152.

Pickering, M. J., & Gambi, C. (2018). Predicting while comprehending language: A theory and review. *Psychological Bulletin*, *144*(10), 1002.

Polyakov, S., Boatman, E., & Thomas, S. (2010). Noun project. In *Noun Project: Free Icons &amp; Stock Photos for Everything*. https://thenounproject.com/

R Core Team. (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org

Roll, M., Horne, M., & Lindgren, M. (2010). Word accents and morphology—ERPs of Swedish word processing. *Brain research*, *1330*, 114–123.

Ryskin, R., & Nieuwland, M. S. (2023). Prediction during language comprehension: What is next? *Trends in Cognitive Sciences*.

Säfken, B., Rügamer, D., Kneib, T., & Greven, S. (2021). Conditional model selection in mixed-effects models with cAIC4. *Journal of Statistical Software*, *99*(8), 1–30. https://doi.org/10.18637/jss.v099.i08

Salverda, A. P., Kleinschmidt, D., & Tanenhaus, M. K. (2014). Immediate effects of anticipatory coarticulation in spoken-word recognition. *Journal of Memory and Language*, *71*(1), 145–163.

Schneider, W., Eschman, A., & Zuccolotto, A. (2012). *E-prime reference guide*. Psychology Software Tools, Inc. https://support.pstnet.com

Schreiber, K. E., & McMurray, B. (2019). Listeners can anticipate future segments before they identify the current one. *Attention, Perception, & Psychophysics*, *81*(4), 1147–1166.

Söderström, P., Roll, M., & Horne, M. (2012). Processing morphologically conditioned word accents. *The Mental Lexicon*, *7*(1), 77–89.

Stone, K. (2021). Bayesian divergence point analysis of visual world data. In *Kate Stone | University of Potsdam*.

https://stonekate.github.io/blog/bpda/

Stone, K., Lago, S., & Schad, D. J. (2021). Divergence point analyses of visual world data: Applications to bilingual research. *Bilingualism: Language and Cognition*, *24*(5), 833–841.

Stone, K., Veríssimo, J., Schad, D. J., Oltrogge, E., Vasishth, S., & Lago, S. (2021). The interaction of grammatically distinct agreement dependencies in predictive processing. *Language, Cognition and Neuroscience*, *36*(9), 1159–1179.

Stowe, L. A. (1986). Parsing WH-constructions: Evidence for on-line gap location. *Language and Cognitive Processes*, *1*(3), 227–245.

Traxler, M. J., & Pickering, M. J. (1996). Plausibility and the processing of unbounded dependencies: An eye-tracking study. *Journal of Memory and Language*, *35*(3), 454–475.

Van Berkum, J. J., Brown, C. M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: Evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(3), 443.

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

Wang, C. B. (2014). *The prosody-syntax interaction in the" yi-bu-qi-ba" rule: A morphologically conditioned tone change in mandarin chinese* [PhD thesis]. The University of North Carolina at Chapel Hill.

Wang, W. S., & Li, K. (1967). Tone 3 in pekinese. *Journal of Speech and Hearing Research*, *10*(3), 629–636.

Weber, A., Braun, B., & Crocker, M. W. (2006). Finding referents in time: Eye-tracking evidence for the role of contrastive accents. *Language and Speech*, *49*(3), 367–392.

Xi, J., Zhang, L., Shu, H., Zhang, Y., & Li, P. (2010). Categorical perception of lexical tones in chinese revealed by mismatch negativity. *Neuroscience*, *170*(1), 223–231.

Zhang, J. (2014). Tones, tonal phonology, and tone sandhi. *The Handbook of Chinese Linguistics*, 443–464.