Listeners do not use Mandarin tone sandhi to make lexical predictions: evidence from printed-word visual world paradigm

Yiling Huo, Effie Lu, Yiming Ning, and Wing-Yee Chow
Division of Psychology and Language Sciences, University College London, London, UK

Comprehenders use a variety of predictive cues to generate predictions about upcoming language input on the fly [1, 2]. Recently, studies have proposed that informative phonological cues, such as Japanese pitch accent and Mandarin Chinese tone sandhi, may be used to predict upcoming lexical items at slow speech rates [3, 4]. In contrast, a study by Huo & Chow [5] suggested that Mandarin Chinese listeners may *not* consistently use tone sandhi in a numeral to predict upcoming words (classifier and noun) in a noun phrase at a naturalistic speech rate. However, in their experiment, the relationship between the tone of numerals and the target nouns was indirect; the numeral's tone was predictive only of the upcoming nominal classifier's tone, not the noun itself. Thus their visual display—line drawings depicting target nouns—may not have provided sufficient support for participants to use numeral tones effectively in identifying target objects. To address this, we utilized a printed-word visual world paradigm to present complete noun phrases, including the numeral, classifier, and noun to the participants in place of line drawings. Two tone sandhi rules—the T3 sandhi and the *yi* sandhi (Table 1)—were tested as in [5]. Eye movement results suggested a non-significant prediction effect. Applying Bayesian principles to divergence point analysis [6] revealed weak or negligible support for a prediction effect. These findings suggest that Mandarin Chinese listeners may not routinely use tone sandhi to anticipate upcoming words, which constitutes a clear exception to the generalisation that comprehenders can use many kinds of cues for generating lexical predictions on the fly.

Methods. Participants (n=35) viewed pairs of written noun phrases on the screen as they listened to non-constraining sentence fragments that ended with one of the written NPs. Each written NP consisted of a numeral, a classifier, and a noun. In the Different Tones (Experimental) condition, the classifier in one of the NPs triggered tone change (tone sandhi) in the numeral. As a result, the numerals in the two NPs are to be realised in different tones, and as such the tone of the numeral served as an early cue for the identity of the target. In the Same Tones (Control) condition, neither classifier triggered tone sandhi, so the numeral's tone was uninformative about the target's identity. If listeners could use the numeral's tone to predict upcoming words, they should be able to direct their eye gaze to the target earlier in the Experimental than the Control condition.

Results. For trials involving the T3 sandhi, divergence point analysis revealed that the onset of looks to the target was 760ms [720, 860] in the Experimental condition and 845ms [840, 900] in the Control condition. For trials involving the *yi* sandhi, this was 663ms [640, 740] in the Experimental condition and 725ms [720, 780] in the Control condition. Estimated differences between conditions (Experiment - Control) were [-160, 0]ms for the T3 sandhi and [-120, 0]ms for the *yi* sandhi, showing a **non-significant trend** of earlier fixations on the target in the Experimental than the Control condition. Applying Bayesian principles to the distribution of divergence points using principled priors revealed **anecdotal to no support for prediction effects** (*yi*: BF10=0.91; T3: BF10=2.42).

Discussion. Our results show that at a naturalistic speech rate, Mandarin Chinese tone sandhi cues do not significantly inform listeners' lexical predictions. This may be because while phonological cues can inform the listener about the phonological form of an upcoming lexical item (e.g. *liang2* → T3), they are not predictive of the semantic or syntactic properties of the word. Consequently, phonological cues may be less effective in lexical predictions as semantic and syntactic cues, which could explain why they are not used as readily. Prediction effects found in previous studies could stem from experiment-specific strategies, driven by artificial presentation methods such as presenting syllables at a fixed rate, which could enhance the salience of relevant phonological cues. Our results suggest that even with a visual display that greatly limited the set of possible upcoming words, it remains challenging to observe lexical predictions based on phonological cues with more naturalistic speech, calling into question the role of phonological cues as an input to lexical predictions.

References

[1] Van Petten, C., & Luka, B. J. (2012). *International Journal of Psychophysiology*, 83(2), 176-190.
[2] Kamide, Y. (2008). *Language and Linguistics Compass*, 2(4), 647-670.
[3] Ito, A., & Hirose, Y. (2024). *Quarterly Journal of Experimental Psychology*.
[4] Shun L, Chen X, Wang S. https://doi.org/10.31219/osf.io/8kdc4
[5] Huo, Y., & Chow, W.Y. (2024). Poster presented at LingO 2024: Oxford Postgraduate Conference in Linguistics, Jesus College, Oxford, UK.
[6] Stone, K., Lago, S., & Schad, D. J. (2021). *Bilingualism: Language and Cognition*, 24(5), 833-841.

Table 1. Illustration of tone sandhi patterns tested in the current study. NB: the true base tone of yi is T1 – how it is pronounced in isolation. Since in the current study, yi never appeared in isolation, we annotated yi4 as the 'base' form as it is compatible with more tones than yi2.

Tone sandhi	Numeral	Base form	Base form example	Sandhi form	Sandhi form example
Yi sandhi	*yi* ('one')	*yi4*	*yi4 zhang1/tiao2/ba3*	*yi2*	*yi2 ge4*
T3 sandhi	*liang* ('two')	*liang3*	*liang3 zhang1/tiao2/ge4*	*liang2*	*liang2 ba3*

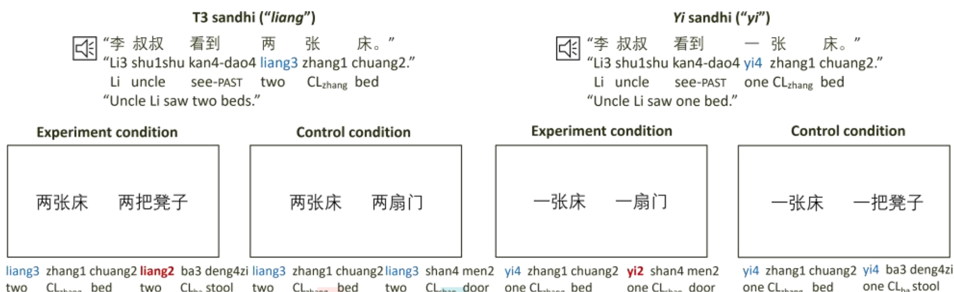

Figure 1. Illustration of stimuli.

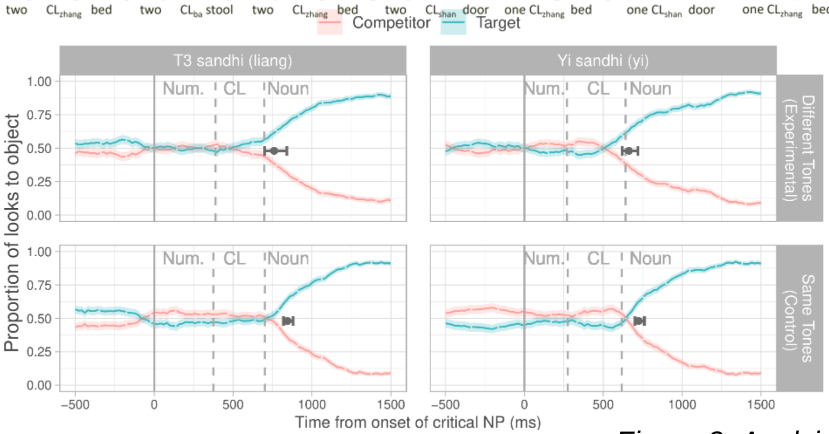

Figure 2. Proportion of looks to the objects. Solid points and error bars represent mean divergence point (onset of looks to target) and 95% confidence interval.

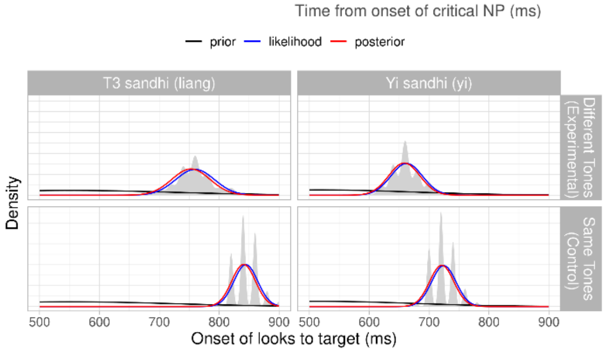

*Figure 3. Applying Bayesian principles to the distribution of divergence points (**onset of fixations on the target**), with principled priors. Prior: Assuming the onset of fixations to the target no earlier than the onset of the numeral, and no later than the offset of the classifier. Likelihood: Distribution of divergence points obtained from the divergence point analysis. Posterior: Posterior distributions calculated using Bayes' theorem.*