2021-12

Anticipatory Tonal Variations Can Facilitate Spoken-Word Recognition: A Visual World Eye-Tracking Study

Jeffrey Jack Green
Brigham Young University, jjgreen@byu.edu

Yan Sun
University of Illinois at Urbana-Champaign

Chilin Shih
University of Illinois at Urbana-Champaign

Original Publication Citation

Sun, Y., Green, J.J., & Shih, C. Anticipatory tonal variations can facilitate spoken-word recognition: A visual world eye-tracking study. 1st International Conference on Tone and Intonation (TAI), Sonderborg, Denmark, December 8. [virtual]

BYU ScholarsArchive Citation

Anticipatory Tonal Variations Can Facilitate Spoken-Word Recognition: A Visual World Eye-Tracking Study

Conference Paper · December 2021

Yan Sun
University of Illinois, Urbana-Champaign

Chilin Shih
University of Illinois, Urbana-Champaign

Anticipatory Tonal Variations Can Facilitate Spoken-Word Recognition: A Visual World Eye-Tracking Study

Yan Sun[1], Jeffrey Green[2] & Chilin Shih[1]
[1]*University of Illinois at Urbana-Champaign*, [2]*Brigham Young University*

The acoustics of speech sounds can vary dramatically depending on phonetic context, yet listeners can reliably process spoken language in a rapid and seemingly effortless manner. Previous research suggests that acoustic variations in speech can actually facilitate, rather than disrupt, the processing of spoken language, and sometimes listeners can even use fine-grained acoustic cues in the unfolding speech signal to predict the sound(s) of an upcoming word. For example, it has been reported that listeners fixate their eyes on target words faster when the pre-target word contains compatible vowel-to-vowel coarticulation cues [1, 2]. Similarly, pitch accents have been found to facilitate discourse comprehension such that listeners "anticipatorily" fixate their eyes on the correct referent when the preceding adjective was assigned the appropriate pitch accent [3]. However, little research has been done to investigate whether anticipatory *tonal* variations can also facilitate spoken-word recognition.

To fill this gap, the current study examines the effect of a specific type of anticipatory tonal variation – third tone sandhi (3TS) in Standard Mandarin – on spoken-word recognition. Standard Mandarin has 4 lexical tones, and 3TS refers to the phenomenon that a low-dipping Tone 3 (T3) changes to a rising tone when preceding another T3. In other words, a canonical T3 is pronounced as sandhi-T3 in anticipation of a following T3. Taking advantage of this anticipatory tonal variation rule, we investigate whether hearing a sandhi-T3 word can facilitate listeners' recognition/prediction of a following T3 word using the "visual world" paradigm [4]. 34 native speakers of Mandarin participated in the study. In each experimental trial, four pictures corresponding to four mono-syllabic words in Mandarin were displayed on a computer screen. One was the target word, one was a competitor, and the other two were distractors. The target word was either T3 or non-T3. The competitor was either: a baseline competitor (BASELINE) which differs from the target word in both tone and segments (T-, S-), a tonal competitor (TONE) which shares the same tone with the target but differs in segments (T+, S-), or a segmental competitor which shares the same segments with the target, but differs in tone (this last condition was included to answer another research question and won't be discussed here). An auditory instruction – *qing3 nin2 dian3 X* ('Please click on *X*') – was played to participants in each trial. Critically, the pre-target word *dian3* ('click on') is pronounced with sandhi-T3 when *X* is a T3 word and with canonical T3 otherwise. Participants' eye movements were monitored during the experiment. Their mouse-clicking responses and reaction time (RT) were also recorded.

The RT data was analyzed using a linear mixed effect regression model. Results show that listeners clicked on the target picture significantly faster when it corresponded to a T3 word than when it corresponded to a non-T3 word under all competitor conditions. This indicates that the sandhi-T3 *dian3* in the carrier sentence did facilitate listeners' processing of a following T3 word.

For the eye-movement data, we calculated fixation proportions on the target and competitor separately for each 50 ms time bin relative to the target word onset. The results are plotted in Figure 1. Panel (a) shows that when the target is in T3, which means the pre-target *dian3* is in sandhi-T3, listeners fixated on the target picture slower in the TONE condition (red contour) compared to the BASELINE condition (green contour) and this difference starts before 200 ms from the target word onset. Panel (b) further reveals that this slower fixation results from increased fixations on the TONE competitor because it is also T3 and thus compatible with the pitch contour of the sandhi-T3 *dian3*. When the target word is non-T3, on the other hand, there are few visible differences in fixation proportions between the TONE and BASELINE conditions because the canonical T3 *dian3* is compatible with both competitor types. Since it is generally agreed in the literature that the time required to program a saccade based on auditory input is about 200 ms [5], these eye-movement patterns suggest that listeners can quickly detect and use 3TS cues in the pre-target word to anticipate an upcoming T3 word before it is encountered. These observations were statistically supported by growth curve analyses which included a third-order (cubic) orthogonal polynomial to capture non-linear changes in fixation proportions over time and participant random effects on all time terms [5] (see Table 1 for detailed results). Therefore, our study provides evidence that anticipatory tonal variations such as 3TS in Standard Mandarin can indeed facilitate spoken-word recognition and listeners can even use such cues to predict the tone of an upcoming word.

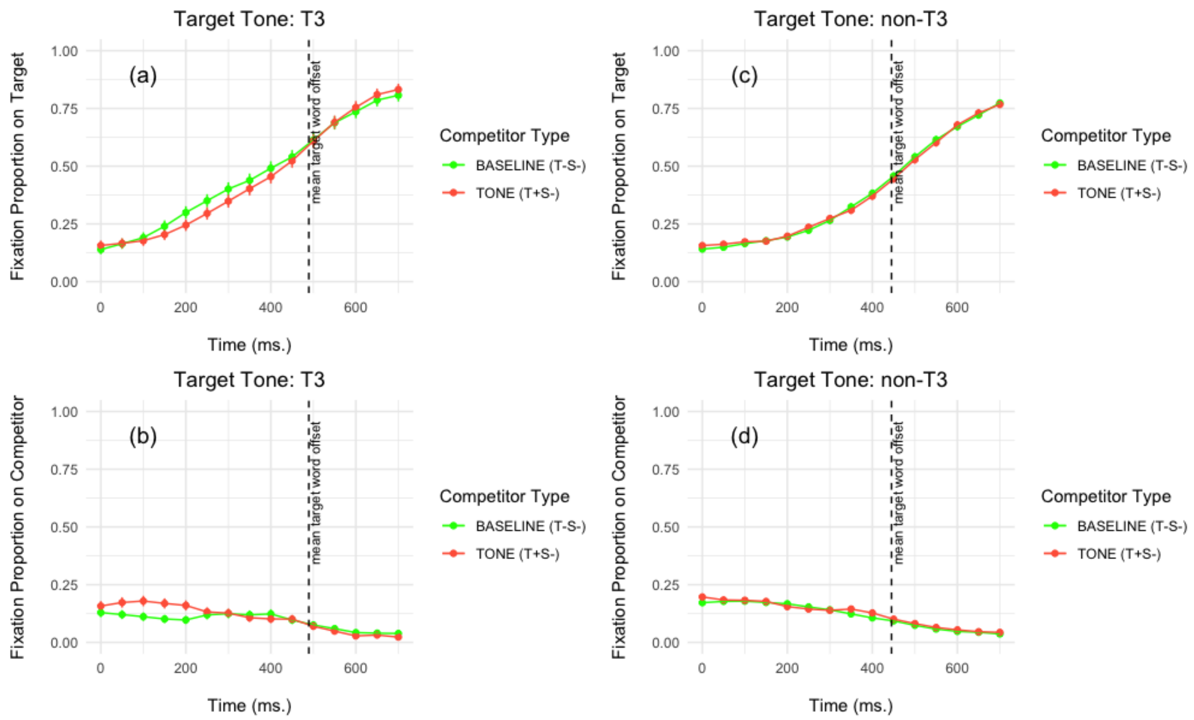

Figure 1. *Fixation proportion on target (top) and competitor (bottom) when target word is T3 (left) and non-T3 (right). The 0ms shows target word onset. The dashed vertical line (y = 489 ms when target is T3 and y = 445ms when target is non-T3) indicates the mean target word offset.*

Table 1. *Growth curve analyses results*

		Target Tone: T3				Target Tone: non-T3			
		Est.	SE	*t*	*p* <	Est.	SE	*t*	*p* <
Fixation Proportion on Target	(Intercept) BASELINE	0.46	0.02	21.34	0.001	0.39	0.02	22.12	0.001
	conditionTONE	-0.02	0.02	-1.22	n.s.	-0.00	-0.01	-0.06	n.s.
	poly1: conditionTONE	0.03	0.03	1.00	n.s.	-0.02	-0.02	-1.59	n.s.
	poly2: conditionTONE	0.08	0.03	2.70	**0.01**	0.02	0.02	1.33	n.s.
	poly3: conditionTONE	-0.05	0.03	-1.90	n.s.	0.01	0.02	0.49	n.s.
Fixation Proportion on Competitor	(Intercept) BASELINE	0.09	0.01	7.52	0.001	0.12	0.01	11.38	0.001
	conditionTONE	0.02	0.01	1.20	n.s.	0.00	0.01	0.69	n.s.
	poly1: conditionTONE	-0.09	0.02	-4.62	**0.001**	0.01	0.01	0.46	n.s.
	poly2: conditionTONE	0.02	0.02	0.98	n.s.	0.01	0.01	0.49	n.s.
	poly3: conditionTONE	0.04	0.02	1.99	**0.05**	-0.02	0.01	-4.72	**0.001**

References

[1] 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.
[2] Tobin, S., Cho, P. W., Jennet, P., & Magnuson, J. 2010. Effects of anticipatory coarticulation on lexical access. In *Proceedings of the Annual Meeting of the Cognitive Science Society,* 32(32), 2200-2205.
[3] 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.
[4] Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. 1995. Integration of visual and linguistic information in spoken language comprehension. *Science*, 268(5217), 1632-1634.
[5] Fischer, B. 1992. Saccadic reaction time: Implications for reading, dyslexia, and visual cognition. In *Eye Movements and Visual Cognition*, 31-45. Springer, New York, NY.
[6] Mirman, D. (2014). *Growth Curve Analysis and Visualization Using R*. Chapman and Hall/CRC.