Virtual-World eye-tracking: The efficacy of replicating word processing effects remotely

Zoe Ovans, Jared Novick & Yi Ting Huang (University of Maryland)

Psycholinguistic research has generated detailed models of moment-to-moment language processing, and has increasingly turned to virtual methods that recruit diverse participants, yield large sample sizes, and remain pandemic-proof. However, there is substantial uncertainty about the feasibility and sensitivity of measurements from remote settings. While methods such as self-paced reading and acceptability judgements replicate well online [1], it is unknown whether fine-grained effects (e.g., within word recognition) will be observable. Recent attempts using visual-world eye-tracking have relied on automatic gaze-detection (e.g., [2,3]), but this requires calibration and can have limited accuracy. To validate the efficacy of remote eye-tracking for word processing, the present study employed a novel webcam paradigm (via *PCIbex* [4]) to semi-replicate Experiment 1 in Allopenna et al., 1998 [5]. This landmark study (cited over 1600 times) revealed listeners' incremental activation of phonemic competitors during spoken-word recognition. It is an ideal candidate for validating remote testing, since real-time fixations track the extent to which subtle acoustic changes incrementally alter predictions of word identity. It has been replicated in laboratory settings (e.g., [6]), but, to our knowledge, not remotely.

Compared to eye-tracking in the lab, webcam eye-tracking introduces additional variability, including participants' screen size, camera quality, internet bandwidth, and environmental distractions. It was our aim to determine whether these factors limit sensitivity to the time-course of word recognition. While data collection is ongoing, 34 participants have been collected from Amazon Mechanical Turk and the university study pool. Some participants had hardware difficulties or did not yield suitable data, but our overall data-retention rate was 79%. We showed listeners an image of a spoken target (e.g., "beaker"), phonological cohort competitors (e.g., "beetle"), rhyme competitors (e.g., "speaker"), and unrelated distractors (e.g., "carriage"). If incremental word recognition is observable in this format, we expect to see looks to the target and cohort-competitor images after word onset, and to a lesser extent, to the rhyme-competitor after word offset. To increase the feasibility of virtual testing, we included only partial-set trials (e.g., with two unrelated objects, target and cohort-competitor) in a Latin square design, reducing the trial number from 96 to 18. This ensured that cohort and rhyme competitor looks were independent, encouraged participants to stay engaged, and reduced video upload time. Looks were recorded through participants' webcams and hand-coded frame-by-frame [7].

As Fig.1 shows, looks to the target increased following disambiguation, confirming that participants successfully link the audio to our visual displays. Looks to the target object began 400ms after word onset, about 200ms slower than lab-based eye-tracking [5]. To examine the extent of competition, we averaged fixations in a 1000ms time-window after word onset, and compared competitor fixations to unrelated controls. As predicted, participants looked to cohort and, to a lesser extent, rhyme competitors after target word offset (Fig. 1). Mixed-effects models reveal more looks to cohort than rhyme and unrelated competitors (ps<.01), though looks to rhymes did not differ from unrelated items (p=.29). Next, we calculated the relative target and competitor frequency and included this as a fixed effect (Fig. 2). Consistent with [8], we found an interaction between frequency and rhyme looks. Participants looked to rhymes more than unrelated controls when rhymes were more frequent than targets (p=.02). Together, this shows that incremental word processing and subtle frequency effects are observable in virtual testing. We conclude that webcam eye-tracking produces similar results to in-lab testing, but eye-movements are slower, and subtle effects like rhyme competition may be harder to detect. Even so, the presence of cohort competition and frequency modulation provides evidence for this method's sensitivity to incremental processing, and provides validation for a new, virtual avenue for visual-world sentence processing research for closely time-locked effects.

Figure 1: Proportion of looks to items surrounding target word onset

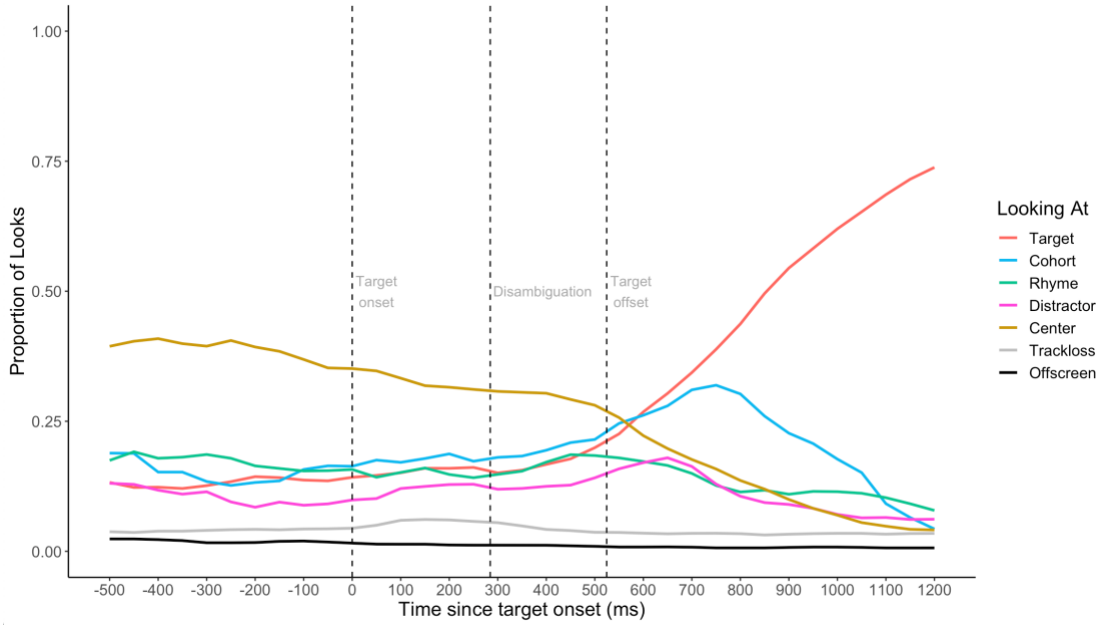
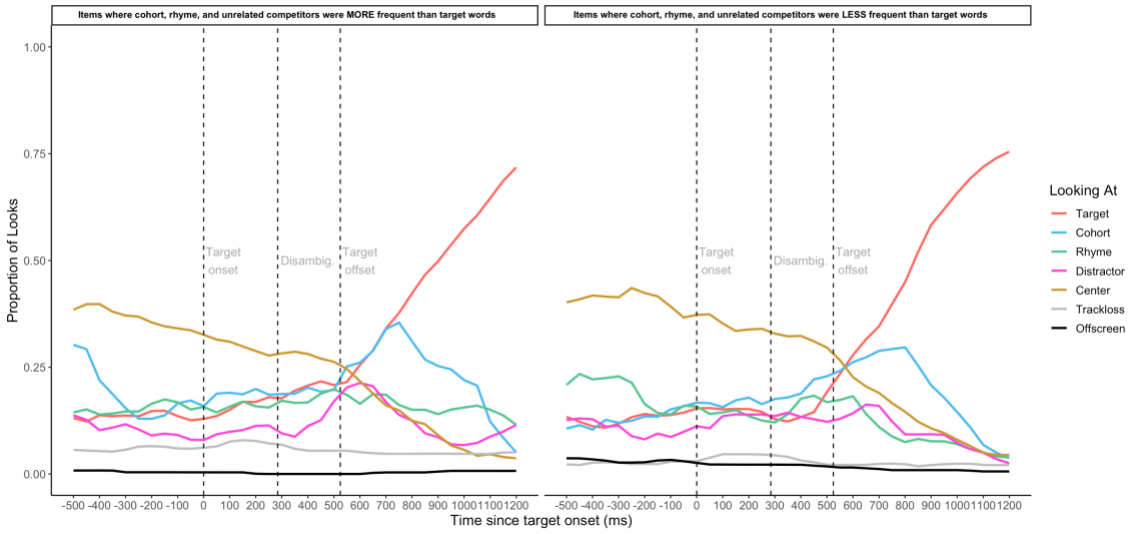


Figure 2a: Items where cohort, rhyme, and unrelated competitors were MORE frequent words than target words

Figure 2b: Items where cohort, rhyme, and unrelated competitors were LESS frequent words than target words



References:

[1] Sprouse, J. (2011). A validation of Amazon Mechanical Turk for the collection of acceptability judgments in linguistic theory. *Behavior research methods*, *43*(1), 155-167.

[2] Xu, P., Ehinger, K. A., Zhang, Y., Finkelstein, A., Kulkarni, S. R., & Xiao, J. (2015). Turkergaze: Crowdsourcing saliency with webcam based eye tracking. *arXiv preprint arXiv:1504.06755*.

[3] Semmelmann, K., & Weigelt, S. (2018). Online webcam-based eye tracking in cognitive science: A first look. *Behavior Research Methods*, *50*(2), 451-465.

[4] Zehr, J., & Schwarz, F. (2018). PennController for Internet Based Experiments (IBEX).

[5] Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of memory and language*, *38*(4), 419-439.

[6] Farris-Trimble, A., & McMurray, B. (2013). Test–retest reliability of eye tracking in the Visual World Paradigm for the study of real-time spoken word recognition. Journal of Speech, Language, and Hearing Research.

[7] Snedeker, J., & Trueswell, J. C. (2004). The developing constraints on parsing decisions: The role of lexical-biases and referential scenes in child and adult sentence processing. Cognitive psychology, 49(3), 238-299.

[8] Dahan, D., Magnuson, J. S., & Tanenhaus, M. K. (2001). Time course of frequency effects in spoken-word recognition: Evidence from eye movements. Cognitive psychology, 42(4), 317-367.