

# Clear speech benefit for word segmentation is modulated by contextual-semantic cues: Evidence from eye-tracking

Zhe-chen Guo<sup>1</sup> and Rajka Smiljanic<sup>2</sup>

<sup>1</sup>Department of Communication Sciences and Disorders, Northwestern University; <sup>2</sup>Department of Linguistics, University of Texas at Austin

## 1. Introduction

- A crucial task in spoken language comprehension is to **segment speech into words** and **resolve lexical competition** (e.g., *two lips* or *tulips*?).
- Listeners use various signal-dependent cues: preboundary vowel lengthening [1, 2], lexical stress [3], etc.
- Listeners also use relatively **signal-independent semantic cues** from the context [4–8].
- E.g., *yellow* is recognized more accurately in *The color of a lemon is yellow* than in *Mom thinks that it is yellow*.
- Lexico-semantic cues dominate signal-dependent cues during word segmentation at least in quiet [9, 10].
- Clear speech (CS)** [11, 12], an intelligibility-enhancing listener-oriented speaking style, improves word segmentation and reduces lexical competition during real-time speech processing in both quiet and in noise [13].
- Due to enhanced acoustic-phonetic cues related to word boundaries.

### Research questions:

- Does the clear speech benefit for real-time word segmentation interact with contextual-semantic cues (“congruent” vs. “incongruent” context; see below)?
- If yes, what would the clear speech effect and its time course be like in the two semantic contexts?

## 2. Methods

### Visual-world Eye-tracking Experiment [14]

- 26 sentences with temporary ambiguity between the **target** (e.g., *ham*) and a **competitor** (e.g., *hamster*) across a word boundary

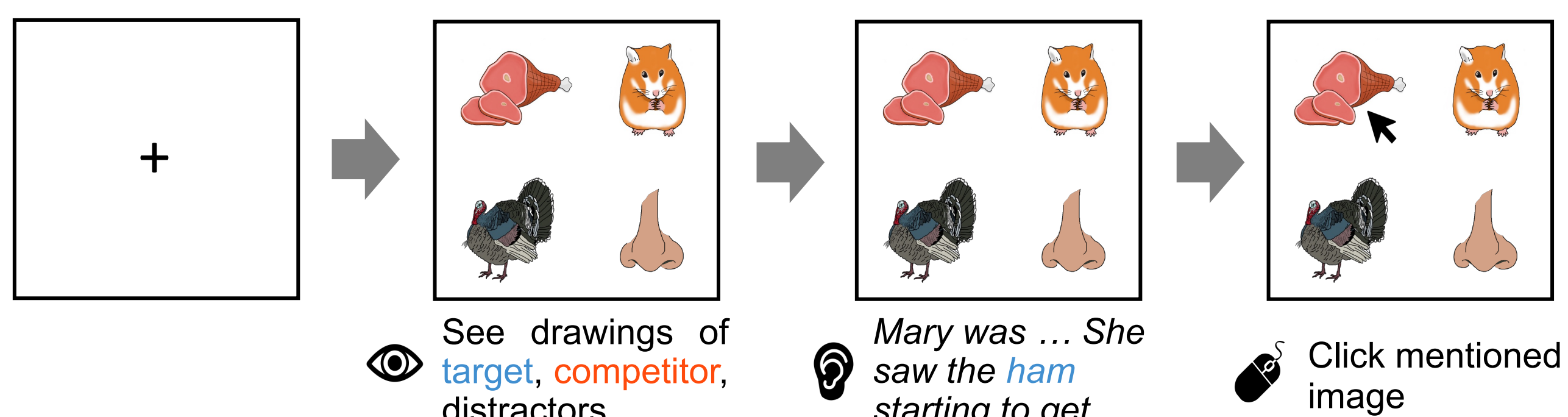
She saw the ham starting to get crispy and brown  
 -----> Disambiguation point (DP)

- Preceded by either a congruent (biasing toward target) or incongruent (biasing toward competitor) context sentence

**Congruent:** Mary was in front of a pork stand.

**Incongruent:** Mary was in front of a pet shop.

- Produced clearly and conversationally by a female native American English speaker

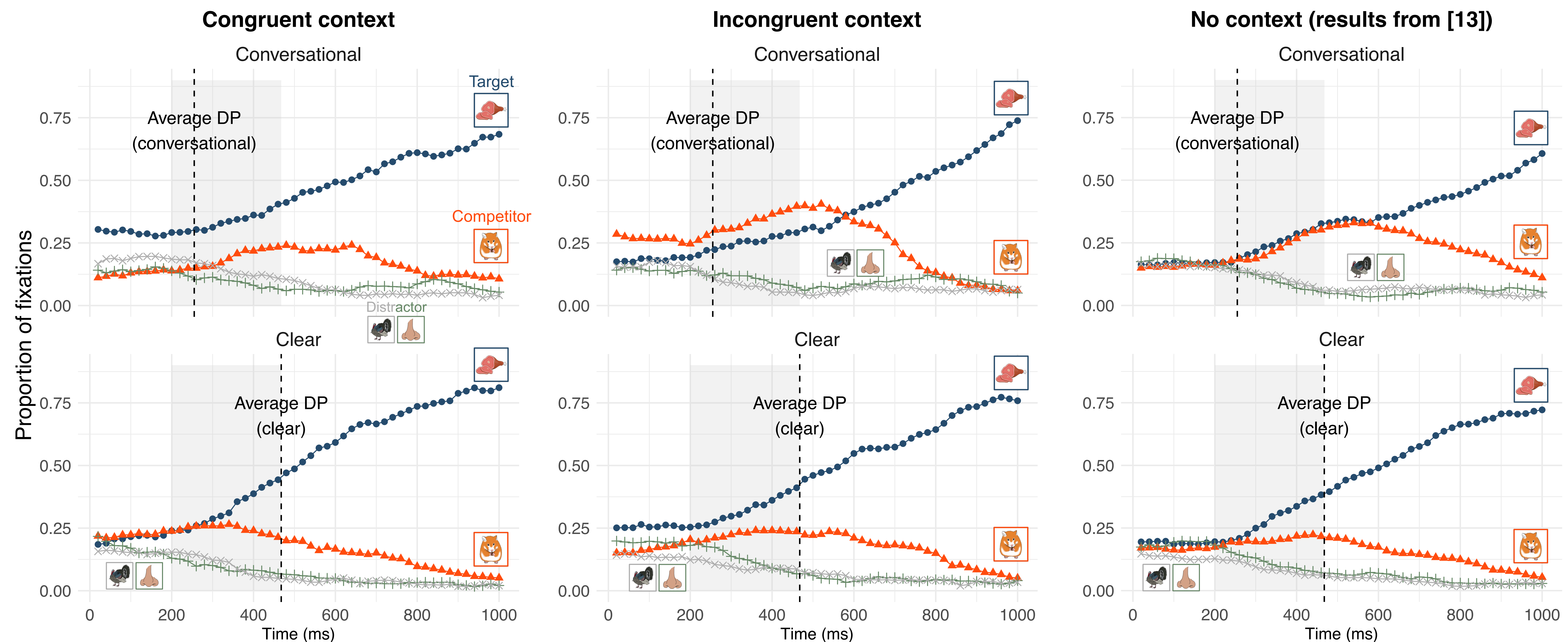


- Recorded right eye at 500 Hz with EyeLink Portable Duo
- 43 native speakers of American English
- Eye fixation to an image reflects the word being considered by the listener as the speech signal unfolds.

## 3. Results

- Analysis time window: 200 ms after target onset to the DP in each clearly spoken sentence (same window was used for its conversational counterpart, thus including some post-DP segments).
- Generalized additive mixed models [15] were fitted to “target advantage” over time:  $EL(\text{prop. of target fixations}) - EL(\text{prop. of competitor fixations})$ , where  $EL$  = empirical logit transformation [16].

Fig. 1. (Left and mid) mean proportions of fixations to the target, competitor, and distractor images in 20-ms time bins within one second after target word onset in the congruent and incongruent conditions. The dashed line marks the average disambiguation point and the shaded area indicates the average analysis time window. (Right) same plots for the results from [13], for target sentences without the preceding semantic context.



- No significant style difference when the congruent context biased listeners toward the target.
- In the incongruent context, the advantage of the target over the competitor in attracting fixations was significantly ( $p < 0.05$ ) greater in CS than in conversational speech.
- This significance emerged at 89% into the analysis window.

- A significantly greater target advantage in CS than in conversational speech emerged at 70% into the analysis window.

## 4. Discussion

- When the semantic context already provided supporting information for the target word, CS did not additionally improve word segmentation and resolution of the ambiguity.
- Contrary to findings that CS improves target word recognition even in high-predictability sentences, though the CS benefit occurs in noise [4, 8].
- However, when the semantic context provided conflicting cues (i.e., favoring competitor), a CS benefit was found.
- The benefit emerged before disambiguating segments were available (replicating [13]).
- This suggested that enhanced acoustic-phonetic cues in CS were relied upon more.
- The results revealed that the context effect was relatively greater in conversational speech than in CS.
- Listeners have been shown to benefit from sentence context more when recognizing more reduced word forms [17, 18].
- Together, the findings suggest that the balance between signal-dependent and -independent information dynamically varies across speaking styles.
- Future research will investigate how the context-style interaction is modulated by (i) listening condition (quiet vs. noise) and (ii) the listener’s language experience (native vs. non-native).

## References

►[1] Salverda, A. P., Dahan, D., & McQueen, J. M. (2003). The role of prosodic boundaries in the resolution of lexical ambiguity in speech comprehension. *Cognition*, 90(1), 51–89. ►[2] Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language*, 35(4), 606–621. ►[3] Tyler, M. D., & Cutler, A. (2009). Cross-language differences in cue use for speech segmentation. *The Journal of the Acoustical Society of America*, 126(1), 367–376. ►[4] Bradlow, A. R., & Alexander, J. A. (2007). Semantic and phonetic constraints for speech-in-noise recognition by native and non-native listeners. *The Journal of the Acoustical Society of America*, 121(4), 2339–2349. ►[5] Dahan, D., & Tanenhaus, M. K. (2004). Continuous mapping from sound to meaning in spoken-language comprehension: Immediate effects of verb-based thematic constraints. *Journal of Experimental Psychology: Learning Memory and Cognition*, 30(2), 498–513. ►[6] Kato, M., & Baese-Berk, M. M. (2023). The effects of acoustic and semantic enhancements on perception of native and non-native speech. *Language and Speech*. ►[7] van der Feest, S. V. H., Blanco, C. P., & Smiljanic, R. (2019). Influence of speaking style adaptations and semantic context on the time course of word recognition in quiet and in noise. *Journal of Phonetics*, 73, 158–177. ►[8] Smiljanic, R., & Sladen, D. (2013). Acoustic and semantic enhancements for children with cochlear implants. *Journal of Speech, Language, and Hearing Research*, 56(4), 1085–1096. ►[9] Mattys, S. L., White, L., & Melhorn, J. F. (2005). Integration of multiple speech segmentation cues: A hierarchical framework. *Journal of Experimental Psychology: General*, 134(4), 477–500. ►[10] White, L., Mattys, S. L., & Wiget, L. (2012). Segmentation cues in conversational speech: robust semantics and fragile phonotactics. *Frontiers in Psychology*, 3, 1–9. ►[11] Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. In W. J. Levelt & A. Marchal (Eds.), *Speech Production and Speech Modeling* (pp. 403–439). Springer Netherlands. ►[12] Smiljanic, R. (2021). Clear speech perception. In L. C. Nygaard, J. Pardo, D. Pisoni, & R. Remez (Eds.), *The Handbook of Speech Perception* (2nd ed., pp. 177–205). Wiley. ►[13] Guo, Z.-C., & Smiljanic, R. (2023). Clear speech facilitates word segmentation: Evidence from eye-tracking. In R. Skarmir & J. Valin (Eds.), *Proceedings of the 20th International Congress of Phonetic Sciences* (pp. 177–181). Guarant International. ►[14] Tanenhaus, M. K., Spivey-Nowtown, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, 268(5217), 1632–1634. ►[15] Lin, X., & Zhang, D. (1998). Inference in generalized additive mixed models by using smoothing splines. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61(2), 381–400. ►[16] Barr, D. J. (2008). Analyzing “visual world” eye-tracking data using multilevel logistic regression. *Journal of Memory and Language*, 59(4), 457–474. ►[17] Ernestus, M., Baayen, H., & Schreuder, R. (2002). The recognition of reduced word forms. *Brain and Language*, 81(1–3), 162–173. ►[18] Kemps, R., Ernestus, M., Schreuder, R., & Baayen, H. (2004). Processing reduced word forms: The suffix restoration effect. *Brain and Language*, 90(1–3), 117–127.