2pSCa12

# Clear speech improves word segmentation in quiet and in noise: Evidence from visual-world eye-tracking


**Zhe-chen Guo and Rajka Smiljanic**

*Department of Linguistics, University of Texas at Austin*


## 1. Introduction

- **Word segmentation** is a fundamental step in speech processing and comprehension (e.g., *two lips* vs. *tulips*).
  - Listeners also need to resolve **lexical competition** (e.g., hearing *cap-* activates *cap*, *captain, captive*, etc. in the mental lexicon).
  - They can use various signal-dependent and relatively signal-independent cues: word stress [1], preboundary vowel lengthening [2, 3], lexical and semantic knowledge [4, 5], etc.
- One way to enhance intelligibility and aid speech processing is to produce **listener-oriented hyperarticulated clear speaking styles** [6, 7].
  - Longer segment duration, vowel space expansion, larger pitch range, etc.
  - Clear speech improves word recognition in quiet and noise [8, 9], memory for spoken information [10], and segmentation of continuously repeated nonsense words presented in quiet in an artificial language learning task [11].

**Research questions:**

1. Does clear speech affect **segmentation of meaningful words** during real-time speech processing?
2. What is the time course of this effect?
3. How is the clear speaking style benefit, if any, modulated by **noise**?

## 2. Methods

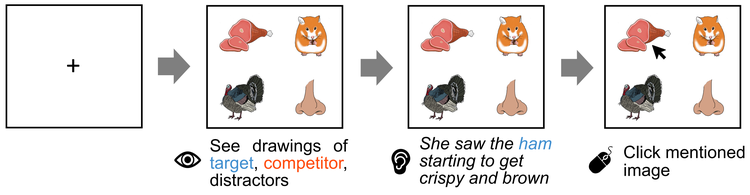
### Visual-world Eye-tracking Experiment [12]

- 26 sentences in which the target word (e.g., *ham*) was temporarily ambiguous with a competitor (e.g., *hamster*) across a word boundary

She saw the ham starting to get crispy and brown

Disambiguation point (DP)

- Produced clearly and conversationally by a female native American English speaker; presented in quiet or in speech-shaped noise at +3 dB SNR



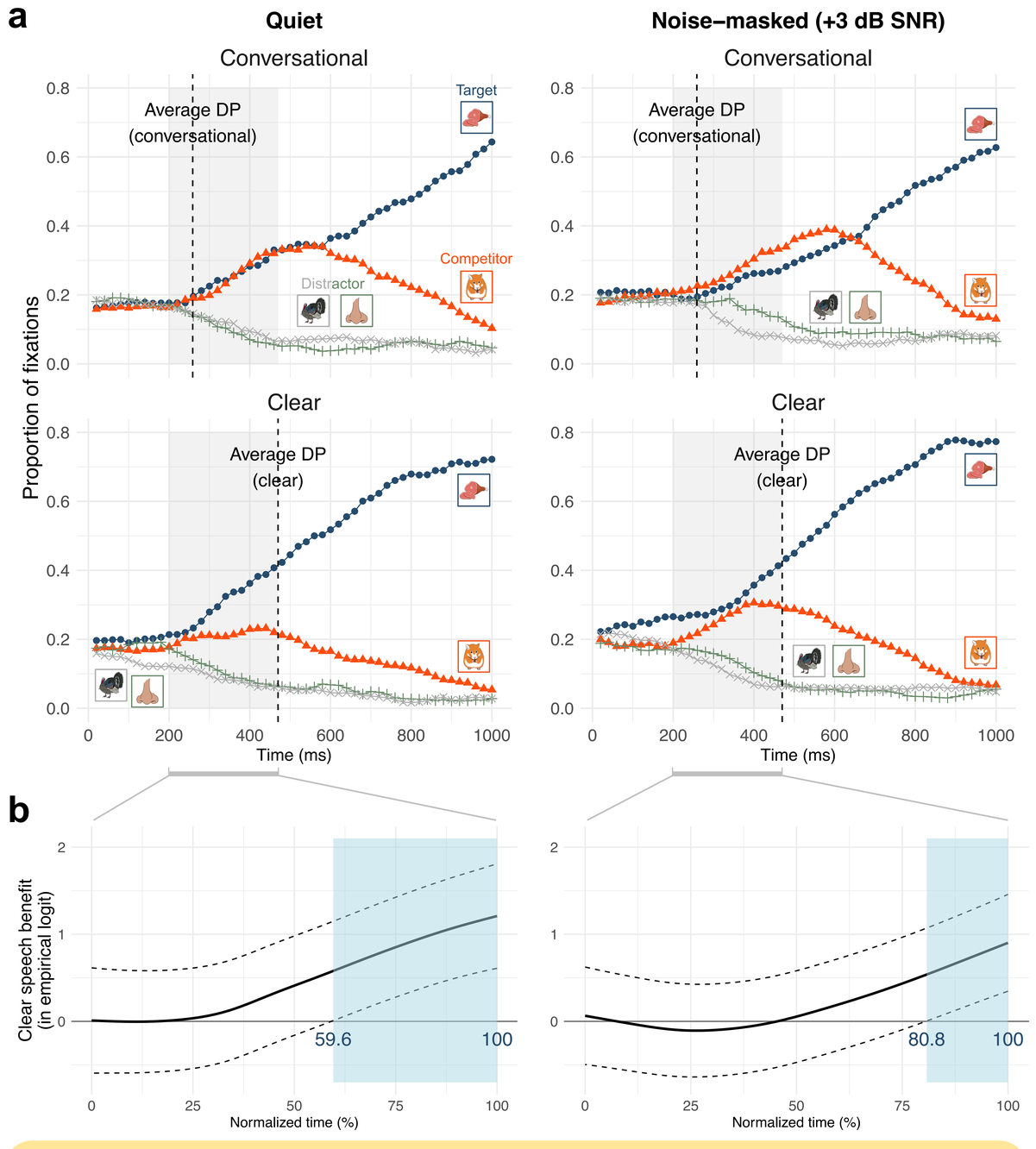

- Recorded right eye positions at 500 Hz with EyeLink Portable Duo
- 76 native speakers of American English

### *Data Analysis*

- Eye fixation to an image reflects what is being considered by the listener as the speech signal unfolds.
- Analysis time window: 200 ms after target word onset to the DP in each clearly spoken sentence (the same window was used for its conversational counterpart, thus including segments after the DP for conversational sentence).
- Fixation proportions were converted to empirical logits [13] and analyzed using generalized additive mixed modeling [14]

## 3. Results

**Fig 1**. (**a**) Average proportions of fixations to the target, competitor, and distractor images in 20-ms time bins within one second after the onset of the target word in quiet and noise listening conditions. The dashed lines mark the average disambiguation points in clear and conversational speech and the shaded area marks the average analysis time window. **(b)** Clear speech benefit (clear − conversational) over time within the analysis window in terms of how much the listeners considered the target over the competitor (expressed as empirical logits of fixation proportions). The dashed lines represent the 95% confidence interval.




- Within the analysis time window (from 200 ms after target word onset to the DP in clearly spoken sentence), listeners fixated the target picture significantly more than the competitor picture ($p < 0.05$) in clear speech compared to conversational speech.
  - This was found in quiet and in noise
- The clear speech benefit emerged at 59.6% into the window in quiet and at 80.8% in noise.

## 4. Discussion

- Relative to conversational speech, clear speech facilitated word segmentation, aiding listeners in disambiguating between the targets and competitors.
  - This occurred even before the listeners heard post-DP segments in clear speech.
- Noise delayed, but did not eliminate, the clear speech segmentation benefit.
  - Consistent with previous research showing that noise induces greater lexical competition [15, 16].
- Clear speech benefit found in the artificial language learning task [11] extends to real-world word segmentation in both quiet and noise. Improved word segmentation may in part underlie the well-documented clear speech benefits for word recognition in noise and auditory memory [9, 10].
- The current findings cannot be accounted for by classic spoken word recognition models based on lexical access and phonemic descriptions of speech inputs (e.g., TRACE: [17]).
  - Including information about enhanced signal-dependent word boundary cues that constrain lexical access could improve the models.
- Future work will examine: (1) how hyperarticulated clear speech interacts with contextual-semantic information; (2) the extent to which speaking clearly improves word segmentation for non-native listeners.

## References


➤**[1]** 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. ➤**[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]** Salverda, A. P., Dahan, D., & McQueen, J. M. (2003). The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition*, *90*(1), 51–89. ➤**[4]** 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. ➤**[5]** White, L., Mattys, S. L., & Wiget, L. (2012). Segmentation cues in conversational speech: robust semantics and fragile phonotactics. *Frontiers in Psychology*, *3*, 1–9. ➤**[6]** Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. In W. J. Hardcastle & A. Marchal (Eds.), *Speech Production and Speech Modelling* (pp. 403–439). Springer Netherlands. ➤**[7]** Smiljanić, 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. ➤**[8]** Bradlow, A. R., & Bent, T. (2002). The clear speech effect for non-native listeners. *The Journal of the Acoustical Society of America*, *112*(1), 272–284. ➤**[9]** 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. ➤**[10]** Keerstock, S., & Smiljanić, R. (2019). Clear speech improves listeners' recall. *The Journal of the Acoustical Society of America*, *146*(6), 4604–4610. ➤**[11]** Guo, Z.-C., & Smiljanić, R. (2021). Speaking clearly improves speech segmentation by statistical learning under optimal listening conditions. *Laboratory Phonology*, *12*(1), 14. ➤**[12]** 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. ➤**[13]** Barr, D. J. (2008). Analyzing "visual world" eyetracking data using multilevel logistic regression. *Journal of Memory and Language*, *59*(4), 457–474. ➤**[14]** Lin, X., & Zhang, D. (1999). Inference in generalized additive mixed models by using smoothing splines. Journal of the Royal Statistical Society: Series B (Statistical Methodology), *61*(2), 381–400. ➤**[15]** Ben-David, B. M., Chambers, C. G., Daneman, M., Pichora-Fuller, M. K., Reingold, E. M., & Schneider, B. A. (2011). Effects of aging and noise on real-time spoken word recognition: Evidence from eye movements. *Journal of Speech, Language, and Hearing Research*, *54*(1), 243–262. ➤**[16]** Brouwer, S., & Bradlow, A. R. (2016). The temporal dynamics of spoken word recognition in adverse listening conditions. *Journal of Psycholinguistic Research*, *45*(5), 1151–1160. ➤**[17]** McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*(1), 1–86.