The role of semantic context in spoken word recognition: Prediction and recovery or optimal cue integration?

Wednesday Bushong (University of Hartford; bushong@hartford.edu) and Daniel LaMarche (University at Buffalo; dlamarch@buffalo.edu)

Introduction. Spoken language is a temporally unfolding signal which requires listeners to quickly process large amounts of information incrementally. Listeners use preceding semantic context in a sentence as a cue for interpreting spoken words (e.g., *She wanted to wear the [?]oat* results in more *coat* than *goat* responses; [1, 2]). Semantic context occurring *after* the target of recognition can also affect its perception [3, 4, 5]. Surprisingly, no study has directly tested whether semantic context affects word recognition differentially when it occurs before vs. after the critical word. Preceding semantic context is frequently characterized as providing a constraint on the interpretation of future words, pre-activating likely word candidates (for review, see [2]); in contrast, subsequent semantic context is often seen as a repair mechanism most likely to exhibit effects on perceptually ambiguous stimuli [3]. This view would predict that semantic context should affect word recognition more when it occurs before vs. after the target of recognition. By contrast, theories which view spoken word recognition as optimal cue integration would predict that semantic context should be treated identically regardless of where it occurs [5, 6]. This study aims to estimate whether relative timing affects the use of semantic context in spoken word recognition.

Stimuli. We developed English sentence stimuli in which we manipulate (i) voice-onset time (VOT) of a critical target word with a [b/p] onset, (ii) a contextual cue which biases toward either the /b/ or /p/ interpretation of the target word, and (iii) whether the contextual cue appears before or after the target word. Stimuli thus form quadruplets:

1(a) I don't mind **[bees/peas]**, but I hate **squash** more than anything. (p-biasing, context-after)
1(b) I don't mind **squash**, but I hate **[bees/peas]** more than anything. (p-biasing, context-before)
1(c) I don't mind **[bees/peas]**, but I hate **wasps** more than anything. (b-biasing, context-after)
1(d) I don't mind **wasps**, but I hate **[bees/peas]** more than anything. (b-biasing, context-before)

We constructed 28 quadruplets using three different target word pairs.[1] Table 1 shows the properties of the sentence stimuli. We acoustically manipulated the VOT of the target words to create a continuum ranging from /b/ to /p/. To ensure that our VOT manipulation was successful, we conducted a norming experiment with native English participants (n=20). Participants heard the critical target words in isolation, responding whether the word started with /b/ or /p/. As expected, we found that VOT significantly predicted the likelihood of /p/ responses ($p < .001$).

Results. Participants (n=61) listened to sentences like (1a-d) above and responded whether the target word started with "p" or "b". VOT (7 steps ranging from 10-50ms), context, and timing were manipulated within subjects. We conducted a mixed-effects logistic regression predicting /p/ responses as a function of VOT, context, timing, and their interactions. Figure 1 shows the results. As predicted, we found significant effects of VOT ($\hat{\beta} = 4.26, z = 12.66, p < .001$) and context ($\hat{\beta} = 3.22, z = 12.59, p < .001$), replicating previous work on the use of acoustic and contextual cues in spoken word recognition [1-4]. Critically, there was no significant interaction between context and timing ($\hat{\beta} = .25, z = 1.41, p = .16$).

Conclusions. Our results suggest that listeners combine acoustics and semantic context during spoken word recognition. Critically, they are integrated in the same way regardless of their relative timing. These results provide evidence that listeners optimally combine cues during spoken word recognition (e.g., [6]), and against accounts which claim that top-down information plays a special or privileged role in spoken word recognition.

[1] We additionally constructed 16 sentence quadruplets with critical words containing [l/r] contrasts (e.g., lock/rock), which were included as filler trials in our perception experiment.

Context and timing	Mean context distance (syllables)	Mean context distance (words)	Non-target words with b/p onset
b-biasing/before	4	3.84	0.438
b-biasing/after	4.19	3.84	0.5
p-biasing/before	3.66	3.5	0.469
p-biasing/after	4.25	3.91	0.562

Table 1: Statistics of sentence stimuli. Distance does not include target word (i.e., *...peas, but I hate squash...* is a context distance of 4 words.)

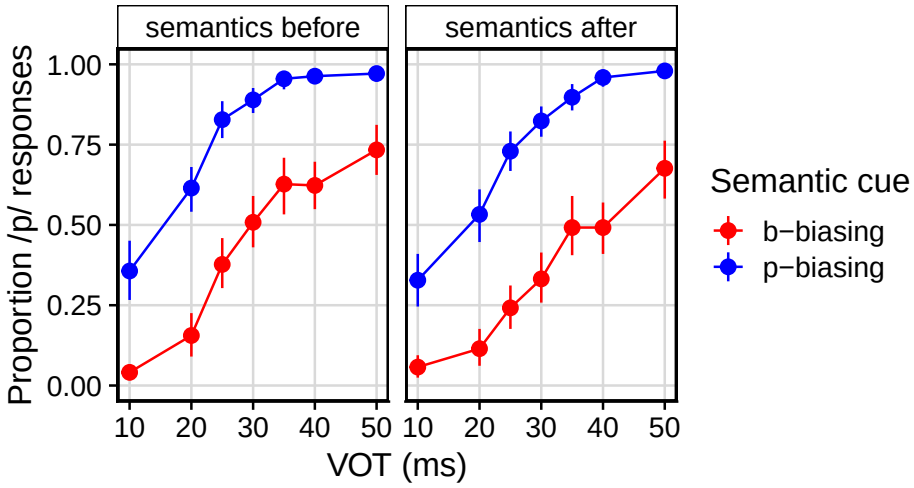

Figure 1: Categorization responses by VOT, context, and context timing. Error bars are 95% confidence intervals.

References

[1] Connine, C. M. (1987). Constraints on interactive processes in auditory word recognition: The role of sentence context. *Journal of Memory and Language, 26*(5), 527-538.

[2] Van Petten, C., Coulson, S., Rubin, S., Plante, E., & Parks, M. (1999). Time course of word identification and semantic integration in spoken language. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*(2), 394-417.

[3] Connine, C. M., Blasko, D. G., & Hall, M. (1991). Effects of subsequent sentence context in auditory word recognition: Temporal and linguistic constrainst. *Journal of Memory and Language, 30*(2), 234-250.

[4] Brown-Schmidt, S., & Toscano, J. C. (2017). Gradient acoustic information induces long-lasting referential uncertainty in short discourses. *Language, Cognition and Neuroscience, 32*(10), 1211-1228.

[5] Bushong, W., & Jaeger, T. F. (2017). Maintenance of Perceptual Information in Speech Perception. In *Proceedings of the Thirty-Ninth Annual Conference of the Cognitive Science Society*.

[6] Norris, D., & McQueen, J. M. (2008). Shortlist B: a Bayesian model of continuous speech recognition. *Psychological Review, 115*(2), 357-395.