Modeling subcategorical information maintenance in spoken word recognition

Wednesday Bushong (University of Hartford; bushong@hartford.edu) and T. Florian Jaeger (University of Rochester; fjaeger@ur.rochester.edu)

Language understanding requires listeners to integrate large amounts of perceptual information before it overwhelms sensory memory. However, cues to linguistic units (sounds, words, etc.) are *distributed* across the speech signal. How then do listeners manage what kinds of subcategorical information they maintain in memory and for how long? Consider the bolded word in Table 1: previous work has found that both the initial acoustic cues on the word itself (e.g., voice onset time) and later context (e.g., forest/fender) affect listeners' interpretation of the word as "tent" or "dent" [1-3,7]. If such evidence reflects subcategorical information maintenance beyond the word boundary, this challenges assumptions in models of word recognition (e.g., [4-6]). Formal models of cue integration across time that can address this question have, however, been lacking. We develop four competing models with different levels of information maintenance, and test them against data like [1-3,7].

Models. We develop four computational models of information maintenance. Figure 1(A-B) displays the formalization and behavioral predictions of each model. The *ideal integration* model assumes that listeners maintain subcategorical information about all cues in the signal over time, and thus optimally integrate those cues (proposed in [7]). The *ambiguity-only* model assumes that listeners are more likely to maintain subcategorical information over time when that information is perceptually ambiguous, and less likely when it is unambiguous (proposed in [1]). The *categorize-discard* model assumes that listeners maintain no subcategorical information about cues over time (proposed in [4]). Finally, we introduce a novel model, *categorize-discard-switch*, which assumes that listeners do not maintain subcategorical information about cues over time, but may change their categorization decisions based on subsequent cues. Notably, several of these models make similar *qualitative* predictions about human behavior, *despite the fact that they make different assumptions about information maintenance*. This highlights the importance of of testing these models quantitatively by fitting them directly to behavioral data.

Experiments. We fit all models to four different behavioral experiments (Ns = 39, 37, 48, 51), three of which are from published sources [7-9]. Participants listened to sentences like those in Table 1 and responded whether they heard "tent" or "dent". Both voice-onset time (VOT) of the target and the bias of the later context were manipulated. **Analysis.** All four models are non-linear mixture models. We implemented hierarchical (mixed-effects) instances of these models using the `brms` package in `R`, and fit them against the data from the behavioral experiments. For each experiment, we measure the performance of the four models as the estimated log predictive density ($elpd_{waic}$)—a measure suitable for non-nested comparison of models with inherently different functional flexibility. **Results.** Figures 1(B) and 2 display model performance for all experiments. In all experiments, the ideal integration model outperformed the ambiguity-only model (Experiments 1, 2, 4: $\Delta elpd_{waic} < 2.5$ SEs $\rightarrow$ "weak" evidence; Experiment 3: $\Delta elpd_{waic} > 5$ SEs $\rightarrow$ "strong" evidence). Further, both the ideal integration and ambiguity-only models strongly outperformed the categorize-discard and categorize-discard-switch models ($\Delta elpd_{waic} > 5$ SEs).

Conclusions. We find consistently strong evidence in favor of the models which posit maintenance of subcategorical information over time. This suggests that listeners are able to maintain subcategorical information about prior linguistic input even beyond the word boundary, in contrast to theories which posit that listeners must immediately discard such information due to memory bottlenecks (e.g., [4-6]). That listeners have much more information available to them over time highlights the need for new theories of speech recognition. This work also demonstrates the importance of formalizing quantitative models of behavior to distinguish between different theories.

References. [1] Connine et al. (1991) *JML* [2] McMurray et al. (2009) *JML* [3] Brown-Schmidt & Toscano

(2017) *LCN* [4] Christiansen & Chater (2016) *BBS* [5] McClelland & Elman (1986) *Cog Psych* [6] Luce & Pisoni (1998) *Ear & Hearing* [7] Bicknell et al. (submitted) [8] Bushong & Jaeger (2017) *CogSci* [9] Bushong & Jaeger 2019 *JASA-EL*

Context	Sentence
Tent-biasing	When the **?ent** in the forest was well camouflaged, ...
Dent-biasing	When the **?ent** in the fender was well camouflaged, ...

Table 1: Example stimuli from Experiments 1-4. "?" indicates a sound along the /t/-/d/ continuum with varying voice onset time (VOT).

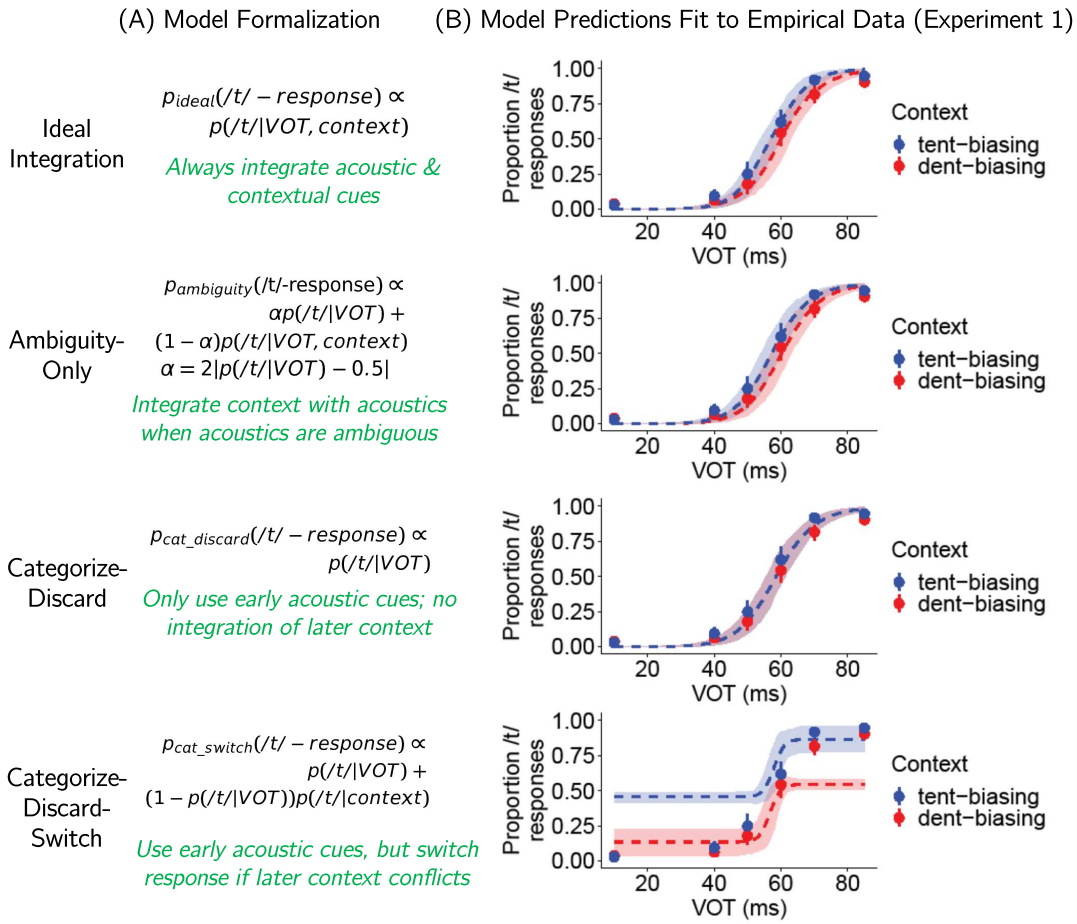

Figure 1: (A): Formalization of each model. (B): Model predictions (dashed lines) fit to empirical data (points; identical across rows). Shaded intervals are 95% confidence intervals.

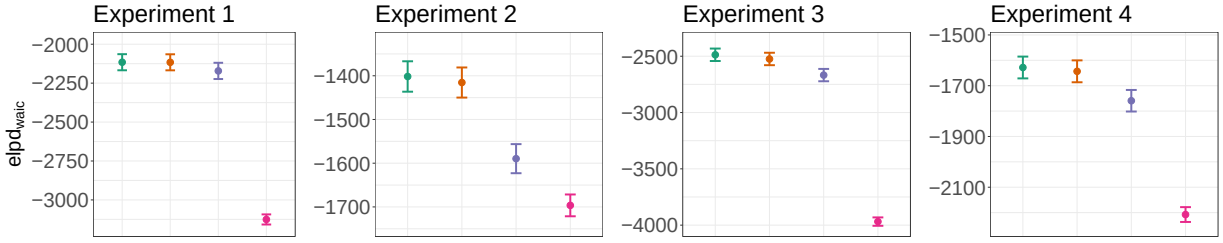

Figure 2: Model fits ($elpd_{waic}$) for Experiments 1-4 (higher values → better fit): ideal integration, ambiguity-only, categorize-discard, categorize-discard-switch.