Modeling Subcategorical Information Maintenance in Speech Perception

Spoken language understanding requires listeners to rapidly and incrementally compress a complex, noisy signal into abstract linguistic units (phonemes, words, etc). A central challenge to speech perception is the asynchronous distribution of relevant information in the speech signal: acoustic cues to a linguistic category can be distributed across the signal. For example, in American English, syllable-onset stop consonant voicing is cued by several acoustic cues present not only on the segment itself (e.g., voice onset time, VOT), but also before and after (e.g., following vowel duration). Optimal integration of these cues would require listeners to retain information about early cues in memory until the later cues are encountered. However, human memory is finite, so that listeners cannot indefinitely retain in memory all details of the signal. Theories of speech perception have thus often adopted some version of the *immediacy assumption*, that listeners do not maintain *subcategorical information* (information beyond the outcome of categorization) about previously encountered segments (e.g., [1-2]). While recent experimental evidence has called this assumption into question [3-6], decisive comparisons of competing theories are still lacking.

The Present Study formalizes—for the first time—multiple computational models of competing hypotheses about subcategorical information maintenance in speech perception. This includes one model that does *not* involve such maintenance but can produce qualitative results that have previously been interpreted as evidence for subcategorical information maintenance. We then conduct a series of perception experiments replicating and extending recent findings of cue integration across time. Finally, we fit the competing models to the results of the experiments, and evaluate which theoretical assumptions best explain human perception.

Quantitative Models. We present several quantitative models of how listeners may maintain subcategorical information, falling into two general classes. Under the first class of models, listeners maintain no subcategorical information after processing a segment. In the most extreme variation of this model, listeners base categorization decisions on only the first-encountered cue. Under a second variation, listeners are allowed to *switch* their categorization decision based on later cues, but critically without referring to the subcategorical information that led to the initial categorization. In contrast to these models, the second general class assumes listeners maintain subcategorical information at least under some conditions. Under one variation, listeners maintain subcategorical information indefinitely, at a level of detail sufficient for optimal integration with later cues. Under a second variation, listeners only maintain subcategorical information when initial cues are perceptually ambiguous.

Perception Experiments. In four web-based perception experiments, we manipulate the VOT of an onset stop consonant of a target word (*tent/dent*), the contextual bias of a later word toward one alternative (*forest/fender*), and the distance between the target and context word (see Table 1). Participants provide a categorization responses for the target word after they hear the entire sentence.

Results. Several findings emerge across experiments. First, we find that both VOT and context affect participants' categorization judgments, suggesting that listeners integrate both of these cues (Fig 1, top panel). Furthermore, this effect does not decrease as distance between the target word and context word increases (at least up to 9 syllables, the longest distance tested). Second, we find that this maintenance behavior can be modulated depending on the statistics of the current environment, such that when maintaining information would be less useful for categorization, participants engage in it less. Critically, when we fit the competing models directly to participants responses, we find that both types of models can explain the qualitative data, but that models which assume listeners maintain subcategorical information significantly outperform models which do not (Fig 1, bottom panel).

Conclusions. The assumption that listeners rapidly discard detailed information about previous input has been largely unchallenged in the literature. The present work shows that listeners can maintain subcategorical information about segments in memory for periods of time extending beyond not only single segments, but several words. This work also highlights the importance of formal quantitative models to distinguish between competing theories of language processing.

References. **[1]** Christiansen & Chater (2016) *BBS* **[2]** Just & Carpenter (19800 *Psych Review* **[3]** Connine et al. (19910 *JML* **[4]** McMurray et al. (20090 *JML* **[5]** Szostak & (Pitt0 2013 *APP* **[6]** Brown-Schmidt & Toscano (2017) *LCN*

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

Table 1: Example stimuli from perception experiments. “?” indicates a sound along the /t/-/d/ continuum with varying voice onset time (VOT).

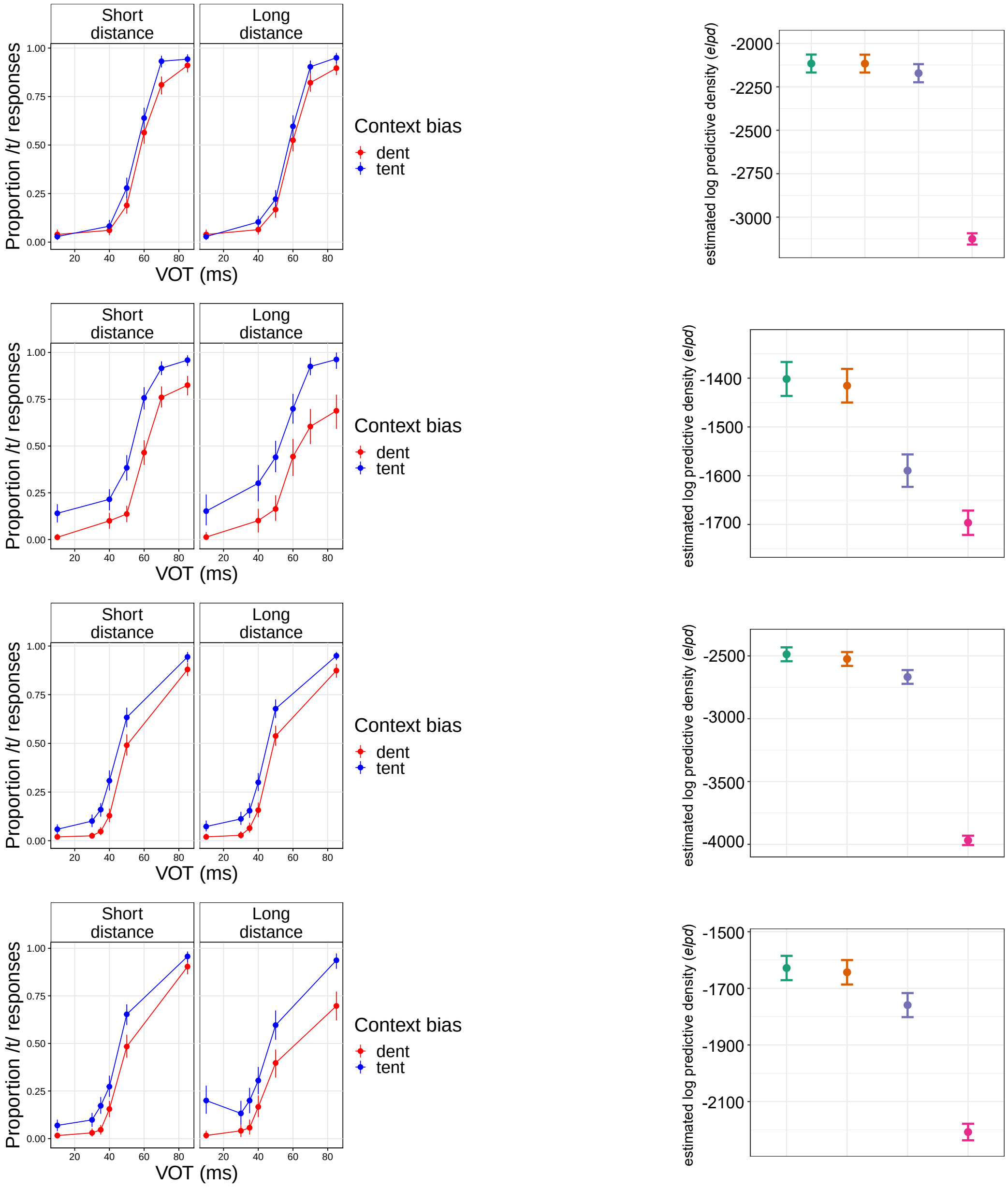

Figure 1: Left panel: exmpirical results for Experiments 1-4. The magnitude of observed effects varies due to variations in experimental manipulations. Right panel: model fits for each experiment. Green and orange represent models assuming subcategorical information maintenance; purple and pink represent models assuming no subcategorical information maintenance.