Cue Reliability and Re-Weighting in Speech Perception

Wednesday Bushong (wbushong@ur.rochester.edu) and T. Florian Jaeger (fjaeger@ur.rochester.edu), Brain & Cognitive Sciences, University of Rochester

During speech perception, listeners integrate bottom-up cues from the signal with top-down cues. There has been much interest in how listeners weight different cues during integration, be they multiple perceptual cues or the integration of perceptual and top-down cues [e.g., 1, 2]. This work has focused on cue integration during processing, and has generally assumed that cue weights are static. However, cue reliability is not constant across situations (e.g., it may vary by talkers, discourse context, etc). Here, we explore whether listeners are sensitive to the relative reliability of cues in the current input, and learn to re-weight cues accordingly. Specifically, we test the hypothesis that the correlations between cues influence cue weighting: the less correlated a cue is with the others, the less reliable it is deemed to be and the more it is down-weighted over time [3, 4]. We investigate how listeners learn to re-weight bottom-up acoustic cues and top-down semantic cues in speech perception when these two cues are either uncorrelated or highly correlated. **Methods.** We present listeners (N=106) with sentences like those in Table 1, and they judge whether they heard "tent" or "dent" (following [5]). Two cues are varied: the acoustics of the sound range from /t/-/d/ (using VOT); and a binary semantic cue biases toward either "tent" or "dent". We divide subjects into two exposure groups differing in the amount of cue conflict they encounter. In the High-Conflict Group, the acoustic cue and semantic cue conflict on half of all trials—i.e., the semantic cue is uncorrelated with the acoustic cue. In the Low-Conflict Group, we reduce the number of conflict trials, increasing the correlation between semantic and acoustic cues (Figure 1).

Semantic Cue	**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. The "?" indicates onset of target word (ranging from /d/ to /t/). Bolded words indicate the semantic cue.

Results. Analysis 1 assessed the overall effect of the two cues using a mixed-effect logistic regression model. We found strong main effects of both VOT ($\beta = 0.1, p < 0.001$) and semantic cue ($\beta = 0.9, p < 0.001$), confirming that both cues affect perception (Figure 2a). We found that the semantic cue effect was much larger in the Low-Conflict Group than in the High-Conflict Group (simple effects significant at all VOTs tested; $\beta s >= 0.43, ps < 0.001$; Figure 2b), suggesting that the High-Conflict listeners relied less on semantic cues. There was an interaction between VOT and group such that the Low-Conflict group's VOT cue weight was higher than the High-Conflict group's ($\beta = 0.014, p < 0.001$). Analysis 2 assessed the weighting of these two cues over time using non-linear modeling (generalized additive mixed models (GAMMs)). We found a significant trial by semantic cue by group interaction, such that semantic cues were down-weighted over time for the High-Conflict group, but not the Low-Conflict group ($\chi^2 = 12.34, p < 0.01$; Figure 3a). We also found a trial by VOT by group interaction such that VOT was down-weighted over time for the High-Conflict Group, but not for the Low-Conflict group ($\chi^2 = 193.6, p < 0.001$; Figure 3b). However, the magnitude of cue down-weighting was larger for semantic cues than acoustic cues. **Conclusions.** Listeners seem to re-weight cues depending on their reliability over time: specifically, listeners who encounter high levels of conflict between acoustic and semantic cues over time down-weight the semantic cues, relying instead primarily on acoustic cues. This effect emerges gradually over time, suggesting that listeners cumulatively track the correlations of cues in their exposure to guide cue re-weighting. These findings highlight the adaptivity of speech perception, and point to implicit learning processes that continuously take advantage of the current statistics of the environment to support optimal processing.

[1] Oden & Massaro (1978) *Psych Review*. [2] Toscano & McMurray (2010) *Cognitive Science*. [3]

Jacobs (2002) *TICS*. [4] Atkins et al. (2001) *Vision Research*. [5] Connine et al. (1991) *JML*.

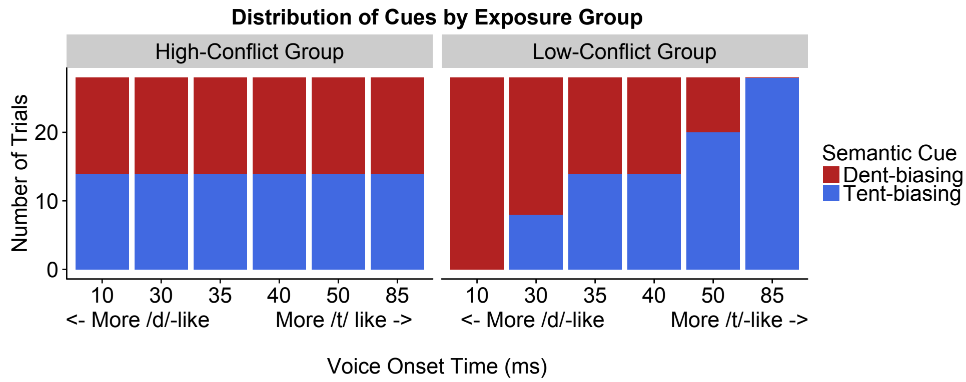

Figure 1: Design of the exposure groups. The High-Conflict Group encounters cue conflict on about half of all trials, whereas the Low-Conflict Group encounters fewer such conflicts.

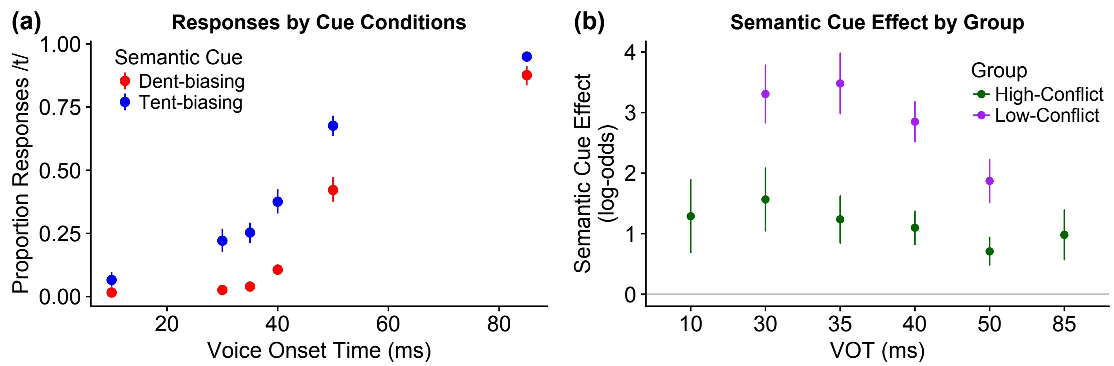

Figure 2: Semantic and acoustic cue effects. **(a)** Proportion responses /t/ by semantic cue, collapsed across exposure group. Error bars are 95% bootstrapped confidence intervals over subject means. **(b)** Weights (coefficients) of the semantic cue in logistic regression are much larger in the Low-Conflict than High-Conflict group at each VOT. Error bars are 95% confidence intervals.

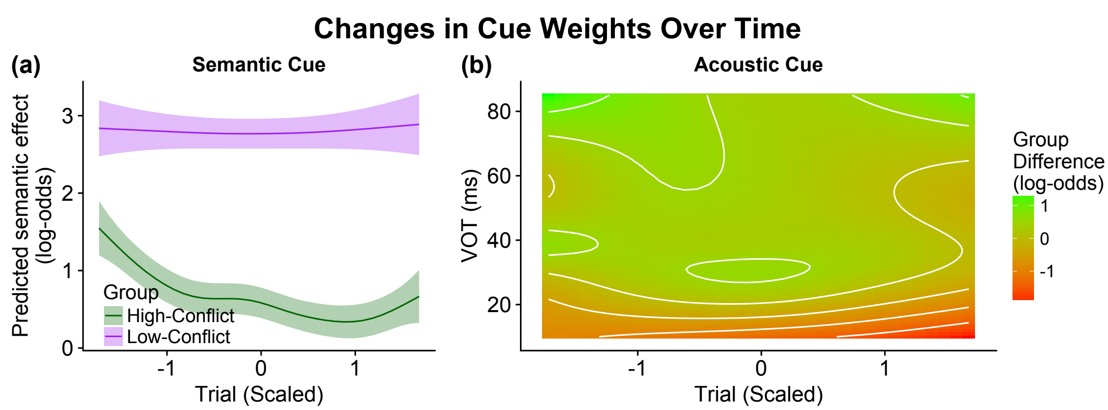

Figure 3: Change in cue weights over trials. **(a)** The High-Conflict group down-weights the semantic cue while the Low Conflict group does not. **(b)** Difference between groups on /t/ categorizations by acoustic cue over time. Lower values on the bottom half of the graph and higher values on the top half of the graph indicate a steeper slope for the Low-Conflict group (i.e., higher cue weight).