

Segmental duration as a cue to sentence structure

Sten Knutsen, Karin Stromswold & Dave Kleinschmidt (Rutgers University)

sten.knutsen@rutgers.edu

Introduction. In order to parse speech in real time, listeners ought to use any informative cues available. Here we investigate the role of segmental duration. On the one hand, previous work has shown that listeners are sensitive to variations in duration, changing their inferences about lexical/syntactic structure when durations are manipulated [1-3]. On the other hand, some production studies have found statistically significant differences in the *mean* durations of analogous segments across different lexical/syntactic structures [1,4-6]. However, a difference in *means* does not necessarily mean that the *distributions* of these durations make individual token durations sufficiently informative to be useful. The goal of the present work is to use production data to quantify how *informative* segmental duration is about syntactic/lexical structure. To this end, we used a Bayesian classifier to model how well a listener could guess the syntactic structure of a temporarily ambiguous sentence in a simulated gating task from natural variation in segmental durations.

Data. We analyzed voice recordings of 8 native English speakers. Each spoke 28 temporarily ambiguous active/passive sentence pairs (Fig. 1), differing only in the choice of verb stem and agent/patient. All sentence pairs were syntactically ambiguous up until the verbal inflection. After excluding 9 tokens due to speaker error, there were a total of 439 recorded sentences. We hand-coded the durations of the onset, nucleus, and coda of the three syllables leading up to the disambiguating verbal inflection (corresponding to the noun, auxiliary, and verb stem).

Model. Our model is based on an ideal listener model, where it is assumed that listeners have implicit knowledge of segmental duration distributions for active and passive sentences. Given these distributions, the model can infer the posterior probability that a particular token belongs to one distribution or the other. We then used an incremental Bayesian belief update model that accumulates evidence from each segment of a particular sentence as it unfolds. This cumulative posterior, combined with a decision rule, is how we modeled listeners' behavior in a gating task with sentences truncated just before the disambiguating verbal inflection.

Procedure. We first estimated active and passive segment durational distributions based on the mean and variance of our training data. We then calculated the posterior probability for all testing data tokens using Bayes' rule (Fig. 2a). We modeled the accumulation of evidence over segments as the cumulative posterior probability, calculated as the cumulative sum of the log-likelihood ratios of segments in each sentence (Fig. 2b). To model behavior in the gating task, we used a winner-take-all decision rule, where the model always guesses the structure with the highest posterior probability (Fig. 2c). To obtain an unbiased estimate of classifier accuracy, we used leave-one-speaker-out cross-validation: holding out one speaker's data for testing while training on the balance of data, and repeating this process for each speaker.

Results. Our classifier output can be seen in Figure 2a, where each blue point represents the posterior of a single token. While for most segments the posteriors are clustered around 0.5 — indicating that they provide little evidence one way or another — the verb stem vowel (e.g., the 'i' in kiss) stands out visually as it carries the largest difference in active/passive distributions. Figure 2b shows the trajectory of evidence accumulation for each sentence in the dataset. The average cumulative posterior probability of each sentence's true structure (Fig. 2b, red line) rises above chance level by permutation test (the red ribbon) at the onset of the verb stem (e.g., the 'k' in kiss) culminating in 0.67 probability by the end of the verb stem, just before the sentence is disambiguated morphosyntactically. Lastly, our modeled winner-take-all accuracy of 74% (Fig. 2c) is in qualitative agreement with previously recorded behavioral results, where participants averaged between 62 and 83% accuracy in gating tasks with analogous sentences truncated before the verbal inflection [4,5].

Conclusion. Our results indicate that there is indeed sufficient information contained in the duration of individual segment tokens so as to be useful to listeners in real-time sentence processing. Whether duration serves as a direct cue to upcoming syntactic structure or indirectly influences such an inference via phonological or morphological levels remains an open question.

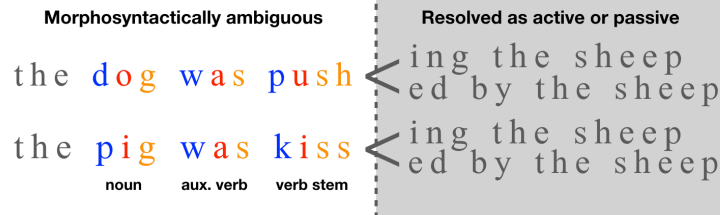


Figure 1: Two sentence pair examples. For each sentence, we coded the **onset**, **nucleus** and **coda** of: the last syllable of the first noun; the auxiliary verb was; the verb stem.

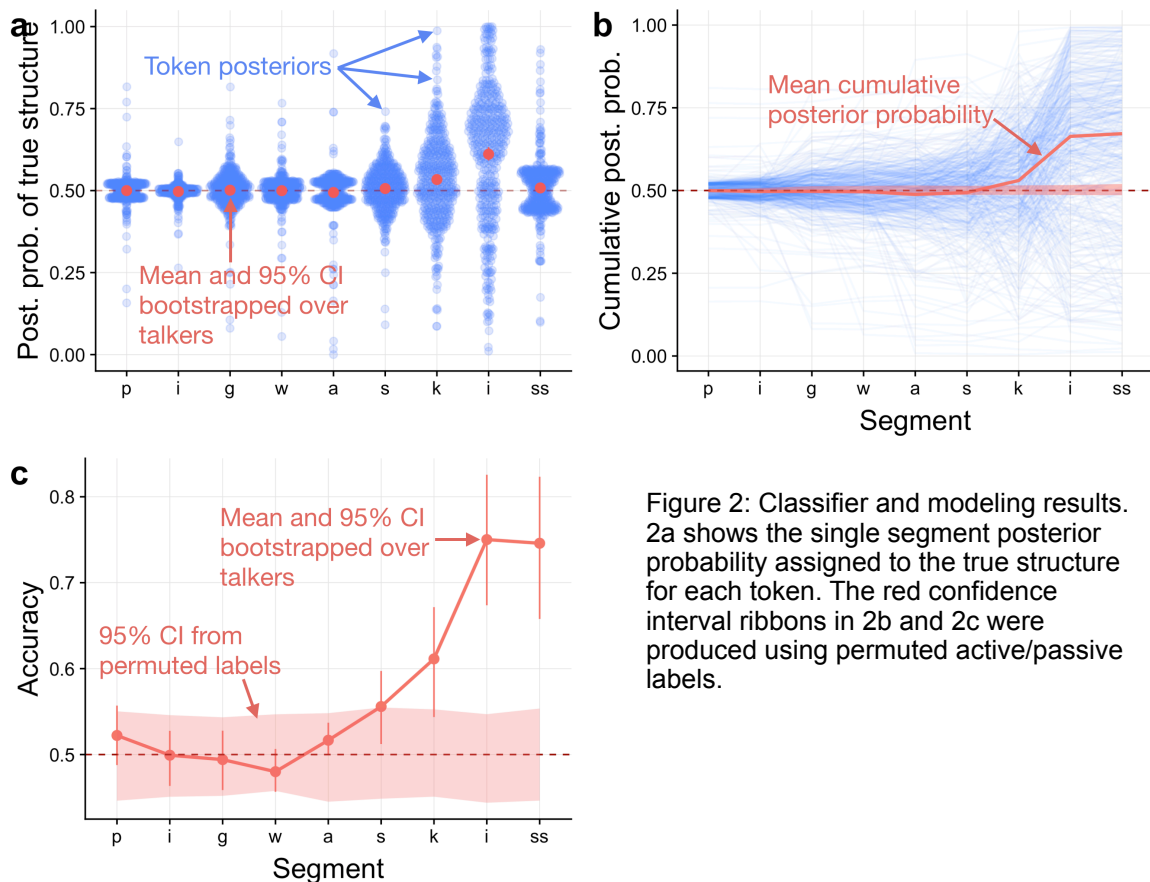


Figure 2: Classifier and modeling results. 2a shows the single segment posterior probability assigned to the true structure for each token. The red confidence interval ribbons in 2b and 2c were produced using permuted active/passive labels.

References

- [1] Beach, C. M. (1991). The interpretation of prosodic patterns at points of syntactic structure ambiguity: Evidence for cue trading relations. *Journal of memory and language*, 30(6):644–663.
- [2] Brown, M., Salverda, A. P., Dille, L. C., & Tanenhaus, M. L. (2011). Expectations from preceding prosody influence segmentation in online sentence processing. *Psychon Bull Rev*, 18:1189.
- [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: 51–89
- [4] Stromswold, K., Lai, M., Rehrig, G., & de Lacy, P. (2016). Passive sentences can be predicted by adults. Poster presented at CUNY 2016
- [5] Stromswold, K., Kharkwal, G., Sorkin, J. E., & Zola, S. (submitted). Tracking the elusive passive: The processing of spoken passives.
- [6] Rehrig, G., Beier, E., Chalmers, E., Schrum, N., & Stromswold, K. (2015). Robust acoustic cues indicate upcoming structure in active and passive sentences. Poster presented at CUNY 2015