Testing three hypotheses of prosodic processing during lexical access

Andrés Buxó-Lugo, University at Buffalo, SUNY (buxolugo@buffalo.edu)

Listeners use prosody to aid their comprehension of words and sentences and to generate predictions about upcoming content (e.g., Dahan, Tanenhaus, & Chambers 2002; Snedeker & Trueswell, 2004; Ito & Speer, 2008). For example, if a word is said with ambiguous segmental properties, then contrastive focus can provide crucial information to differentiate it from similar sounding words (e.g., in "bin – *?IN"* the second word is more likely to be "pin" than "bin"). The exact way that listeners combine prosodic information to arrive at an interpretation of others' speech is, however, unclear. Here we specify three possible processes by which listeners possibly integrate prosodic information with other facets of language comprehension, specifically segmental cues to lexical identity. We then evaluate these models against a behavioral experiment manipulating prosodic content and segmental cues.

To investigate this question, we defined three algorithmic-level accounts (Marr, 1982) of the integration of prosodic and segmental information. Each computational model begins with a logistic function that accounts for listeners' lexical categorizations based on segmental cues (here, voice onset time) with neutral intonation. This baseline function was then tailored to different hypotheses to yield predictions about what responses would look like when the word carried contrastive focus under each hypothesis. The <u>Uniform Prosody Usage Hypothesis</u> states that listeners' use of prosodic information is constant regardless of the ambiguity/clarity of segmental cues. This hypothesis predicts that the effect of prosody on word recognition will be constant regardless of a word's segmental cues. The <u>Two-Stage Prosody Hypothesis</u> states that listeners first try to determine lexical identity from the segmental cues alone but make use of prosody if the segmental cues are ambiguous enough. Under this hypothesis there is a range of segmental cue values that is considered ambiguous, and listeners only use prosody if the segmental cues are within this range. Lastly, the <u>Uncertainty Integration Hypothesis</u> states that listeners weigh prosodic information continuously as a function of the segmental cue ambiguity. Under this hypothesis the effect of prosody is strongest when segmental cues are at the point of greatest ambiguity, and it is gradually scaled down as segmental cues become less ambiguous. Implementations of each of these models can be seen in Table A.

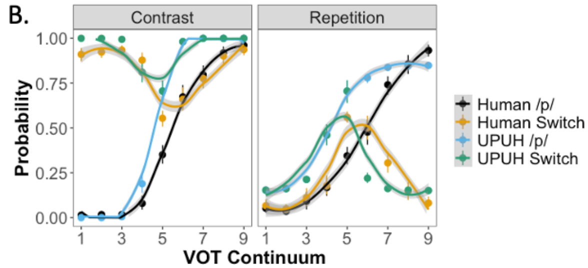
Data from a modified two-alternative forced choice experiment with 61 speakers of American English were used to estimate baseline categorization curves (one per participant) and different combinations of parameters were explored to find models for each hypothesis that best fit with the human data. In this task, participants saw two pictures on a screen and heard two audio recordings per trial. Critical trials consisted of /b/-/p/ minimal pairs, where the first word heard always had unambiguous segmental cues and neutral intonation. After responding to the first word, participants responded to a second word, where we manipulated the onset's segmental cues and whether the word had neutral or contrastive intonation. We were interested in the probabilities of responding that the critical second word started with /p/ and that the second word was judged to be a different word from the preceding word. Models were assessed by calculating Mean Squared Error between the model predictions and the human responses. Results showed that the Uniform Prosody Usage Hypothesis yielded the worst fit to the human data, suggesting that a strictly additive process of integration is not sufficient to capture how humans integrate prosodic and segmental information. The best fit was provided by the Uncertainty Integration Hypothesis, suggesting that listeners weigh the influence of prosody in a continuous manner depending on the uncertainty over the segmental cues. Multilevel logistic regression models supported that responses were a product of an interactive process ($p < .001$).

In conclusion, results from human listeners and from model simulations showed that the integration of prosodic and segmental information is best captured by an interactive process where the influence of prosody is dynamically weighed based on the segmental properties of the input.

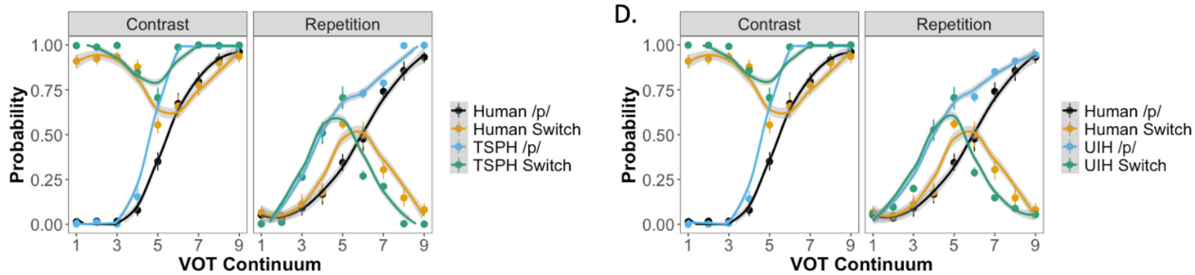
A.

Hypothesis	Alteration to baseline categorization curve
Uniform Prosody Usage Hypothesis (UPUH)	$\begin{cases} f(x) + P \text{ if previous word started with /b/} \\ f(x) - P \text{ if previous word started with /p/} \end{cases}$
Two-Stage Prosody Hypothesis (TSPH)	$\begin{cases} f(x) \text{ if } VOT_c > \textit{threshold} \\ f(x) + P \text{ if } VOT_c \leq \textit{threshold} \text{ if previous word started with /b/} \\ f(x) - P \text{ if } VOT_c \leq \textit{threshold} \text{ if previous word started with /p/} \end{cases}$
Uncertainty Integration Hypothesis (UIH)	$\begin{cases} f(x) + P * \frac{1}{W^{\lvert VOT_c \rvert}} \text{ if previous word started with /b/} \\ f(x) - P * \frac{1}{W^{\lvert VOT_c \rvert}} \text{ if previous word started with /p/} \end{cases}$

B.



C. D.



A. Table showing how each hypothesis was defined. Baseline categorization boundaries (f(x)) were tailored to predict responses in conditions where the critical word had contrastive intonation. Parameters P, threshold, and W were explored to find a combination that resulted in predictions that most closely matched actual human responses. B. Human responses and UPUH predictions for words that had contrastive intonation in cases where the second word was different from the first (Contrast) vs when the second word was the same as the first (Repetition). C. Human responses and TSPH predictions for words that had contrastive intonation in cases where the second word was different from the first (Contrast) vs when the second word was the same as the first (Repetition). D. Human responses and UIH predictions for words that had contrastive intonation in cases where the second word was different from the first (Contrast) vs when the second word was the same as the first (Repetition).

References

Dahan, D., Tanenhaus, M. K., & Chambers, C. G. (2002). Accent and reference resolution in spoken-language comprehension. *Journal of Memory and Language*, *47*(2), 292-314.

Ito, K., & Speer, S. R. (2008). Anticipatory effects of intonation: Eye movements during instructed visual search. *Journal of memory and language*, *58*(2), 541-573.

Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. Freeman.

Snedeker, J., & Trueswell, J. C. (2004). The developing constraints on parsing decisions: The role of lexical-biases and referential scenes in child and adult sentence processing. *Cognitive psychology*, *49*(3), 238-299.