

Visual Grouping and Prosodic Grouping: Effects of Spatial Information on Prosodic Boundary Strength

Edward Holsinger, David Cheng-Huan Li, Elsi Kaiser and Dani Byrd

Department of Linguistics, University of Southern California, Los Angeles, USA {holsinge, lidc}@usc.edu

Introduction

- Prosody refers to the phrasal organization and accentual prominence in speech and may include a wide array of acoustic and articulatory features:
 - The presence/absence/duration of a pause; lengthening of segments and/or articulatory gestures; the pitch contour
- Prosody is also influenced by a number of factors:
 - Constituent structure (Selkirk, 1981; Sanderman & Collier, 1995); speech rate (Fletcher, 1987; Trouvain & Grice, 1999); discourse structure (Ayers, 1994; Venditti & Swerts, 1996)
- Visual input shapes what people talk about, but is there evidence for a more subtle connection?
- Question: **Does visual grouping influence linguistic grouping, as reflected by prosodic boundary strength?**
- To investigate this question, we conducted two experiments:

Exp 1. Production study

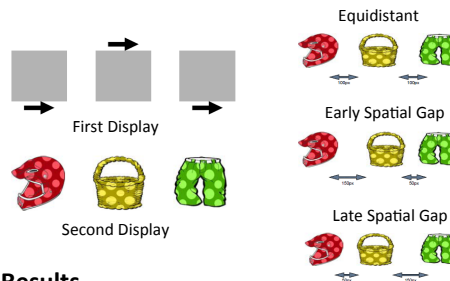
- Analysis = measured pause duration, used this to estimate boundary strength
- Pause duration is easy to recover from acoustic signal
 - But pause duration is only one of the numerous proposed indicators of boundary strength

Exp 2. Perception study

- Analysis = Used listeners' perception of 'connectedness' to estimate boundary strength
- There are *multiple* cues to prosodic boundary strength
 - Using humans as our measurement tool allows us to tap into multiple potential cues of boundary strength
- Existing work has shown that listeners can provide 'connectedness' ratings that relate meaningfully to boundary strength (Krivokapic, 2007)

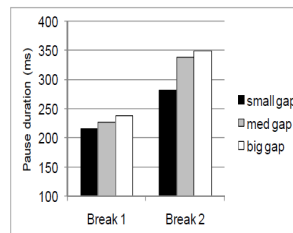
Exp. 1 Production Study

- Participants (n=7) produced scripted utterances based on images on computer screen
- Task: Describe the path of an imaginary little brown mouse as he navigates over or under each object before going into a mouse hole
- E.g. The little brown mouse runs *under the red helmet* {break 1} *over the yellow basket* {break 2} *under the green shorts* and into the mouse hole.
- We manipulated the visual scene layout by changing the distance between the three objects



Results

- Pause duration during break1 and break2 was analyzed with ANOVA (break location X spatial layout)
- Spatially bigger gaps were associated with numerically longer pauses, but not significant



Exp 2. Perception Study

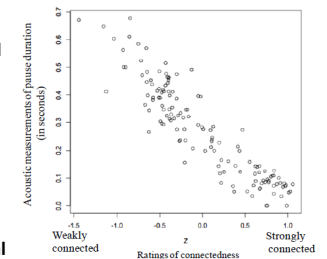
- Listeners' perception of prosodic boundary strength
- Listeners *did not have access* to information about the visual scene
- Participants (n=28) provided ratings of prosodic boundary strength based on what they heard



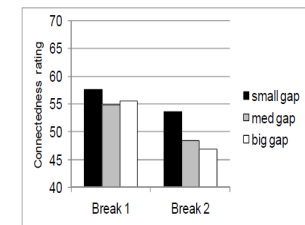
- Task: To rate *how strongly connected* the word of interest is to the word following it, using the slider

Results

- Checking task validity: As expected, connectedness ratings are negatively correlated with pause duration
- Strongly connected = short pause
- weakly connected = long pause
- I.e., Ratings provide meaningful information about prosodic boundaries
- Significant main effect of spatial gap size on connectedness rating ($p < .001$)



- Higher connectedness ratings for small gap than med or big gap (med and big do not differ)
- Importance of **grouping**: Nouns that are grouped together are perceived as more connected (separated by weaker boundaries) than nouns that are excluded from group or ungrouped.



Conclusions

- A cognitively general, non-linguistic property of the visual scene (i.e. group status) influences an abstract property of human speech behavior
- Differences in pause duration between these conditions are very small, making it unlikely that participants could have made these adjustments consciously
- Participants in the perception task did *not* have access to the visual scene information, yet their judgments showed:
 - Highly correlated (in the correct direction) with the measurements collected in the production study
 - Significant differences between conditions which were defined by the visual input used to produce them
- We found effects of spatial grouping on a subtle property of language production: Listeners' rating data indicate that *visual scene layout* influences linguistic production at the level of prosodic boundaries
- By using the fine-grained properties of the human perceptual system, we were able to detect effects of grouping in the visual domain on grouping in the linguistic domain