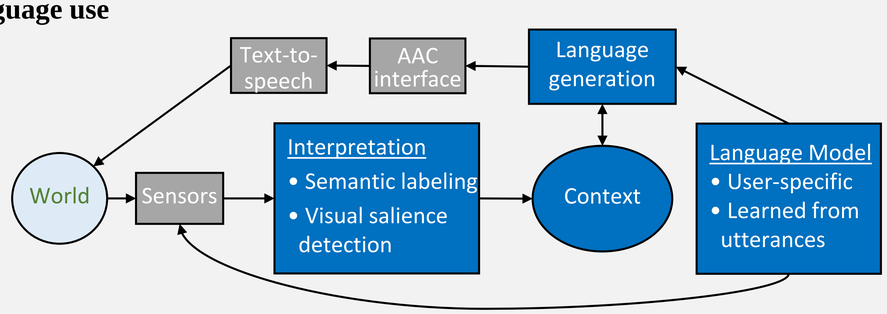
Using Grounded Language for Context-Sensitive Text Prediction

Timothy Lewis, Amy Hurst, Matthew E. Taylor, Cynthia Matuszek
tim22@umbc.edu, amyhurst@umbc.edu, taylorm@eecs.wsu.edu, cmat@umbc.edu

In this work, we focus on using a learned model of language groundings to support language *prediction and generation* – using sensed world state to provide predictions about what a person might wish to say when communicating. A jointly learned model of language and world state is used to interpret sensor data into semantically meaningful *context*, which can then be used to generate contextually informed predictions appropriate to the environment. This approach allows for improved, contextually informed language prediction using existing grounding models.

Goal

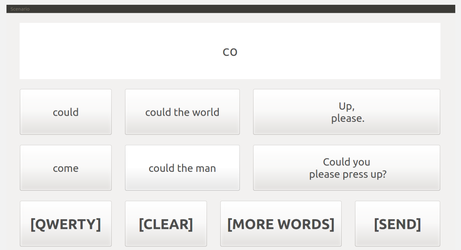
- Develop a predictive text tool driven by a joint **linguistic** and **perceptual** model
- Sense context and apply **semantic labels**
- Generate text predictions for an AAC interface based on **context** plus **model of user's language use**

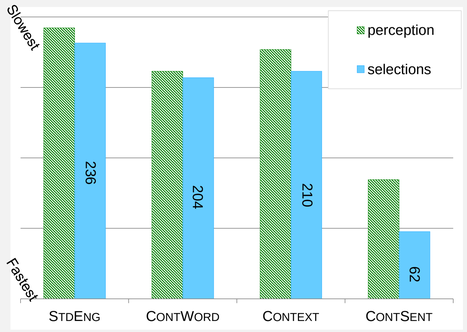


Preliminary Study

"You are on the first floor of a multi-story building. You would like to get to the fifth floor. You are waiting outside the elevator, and you are unable to push the button. However, there is someone standing next to the button panel. What do you say to them?"

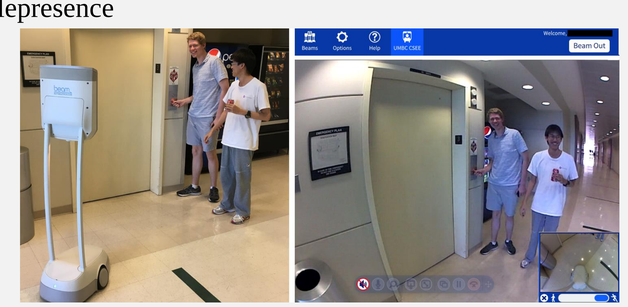
- Users asked to respond to several predetermined scenarios
- Initial language corpus collected via crowdsourcing
- Prototype interface provides users with context relevant text prediction at varying levels
- Quantitative measurements recorded during tasks
- Users provided their qualitative assessment of completing tasks





Approach

- Corpus Collection
  - Crowdsourcing
  - Telepresence



  - End Users
- Context Interpretation
  - Process sensor observations for salient elements in environment
  - Identify elements in the environment paired with speech
  - Learn from non-targeted speech
- Language Learning
  - Implement joint model of language and context
  - Incorporate semantic parsing and perceptual information

$$P(G|x, C), \text{ given data of the form}$$
$$D = \{(x_i, C_i, G_i)|i = 1 \dots n\}$$
$$\rightarrow$$
$$P(x|G, C)$$

Future Work

- Develop more advanced methods to model language and incorporate real world context
- Further experiments with more complete language and grounding models

Cynthia Matuszek et al, 2012. A Joint Model of Language and Perception for Grounded Attribute Learning. In *Proc. of the 2012 International Conference on Machine Learning*. Edinburgh, Scotland, June.
Abhishek Anand et al, 2012. Contextually guided semantic labeling and search for three-dimensional point clouds. *The International Journal of Robotics Research*, p.0278364912461538.