Situational Grounding within Multimodal Simulations

James Pustejovsky and Nikhil Krishnaswamy jamesp@brandeis.edu • nkrishna@brandeis.edu

http://www.voxicon.net
http://github.com/VoxML

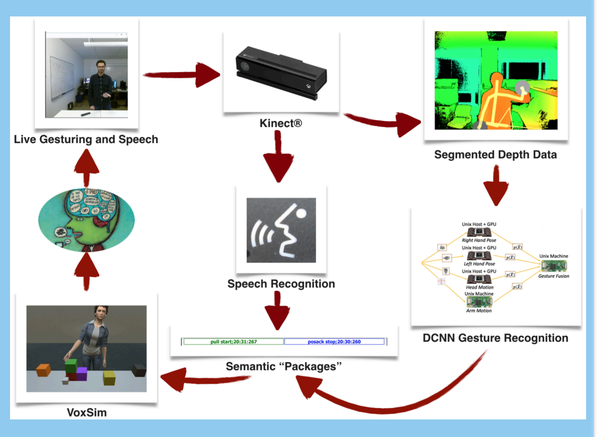
Introduction

3 Definitions of Simulation

1. Computational Simulation Modeling
   - Variables are set, model is run, consequences emerge
   - e.g., climate change, biological pathways, etc.
   - Goal is to arrive at best model using simulation
2. Situated Embodied Simulation
   - User interacts with virtual or simulated world
   - e.g., flight/battle simulator, video games
   - Goal is to simulate agent in situation
3. Embodied Theories of Mind
   - Mental representation of agents and their communicative acts
   - e.g., future or possible outcomes, interpretations of perceptual input
   - Goal is to view semantic interpretation of an expression

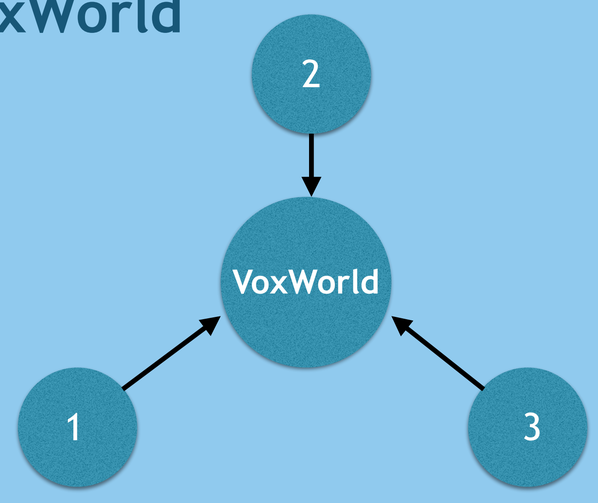
Reasoning in an Interpreted Simulation

VoxSim implementation reasons about consequences of actions taken and needed preconditions



Computer interprets multimodal input — without context of environment, interpreting is intractable

VoxWorld



- Model testing of Computational Simulation Modeling
- Visualized embodiment of Situated Embodied Simulation
- Mode of presentation of Embodied Theories of Mind

Formal Interpretation of Simulations

- Contextualized 3D realization of environment, agents, and salient content of communicative acts, rich semantic typing:
  - Object encoding with action affordances
  - Action encoding as multimodal programs
  - Reveals common ground between parties
- Common ground:
  - Co-situatedness, co-perception, co-attention, co-intent
- VoxML (Visual Object Concept Modeling Language)
  - voxeme : lexeme :: voxicon : lexicon
  - Habitats: situational conditional environment
  - Affordances: behavior driven by structure (Gibsonian) or purpose (telic)

Thanks to Unity!

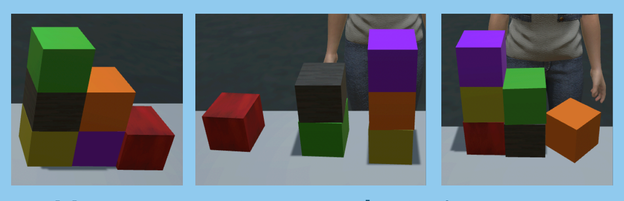
Object model

Action model

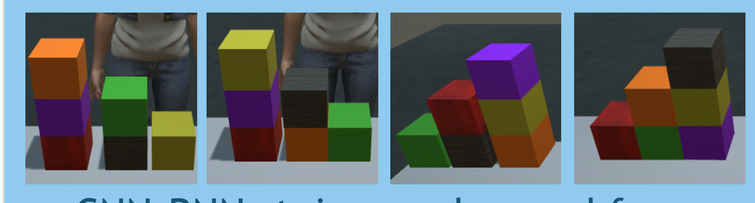
Event model

Decoupled reasoning about objects and actions

Learning by Communication



User-constructed staircases



CNN-RNN staircases learned from qualitative relations

Correcting a sample with multimodal embodied communication

Conclusions

- Deep formal semantics combine with 3D environments to enable "computational embodied cognition"
- Gaming technologies provide powerful platforms to gather data for deep learning and commonsense reasoning
- Game engines do "heavy lifting" of graphics, physics, UI, etc.
- Enable novel research in simulation-based understanding of human and machine intelligence