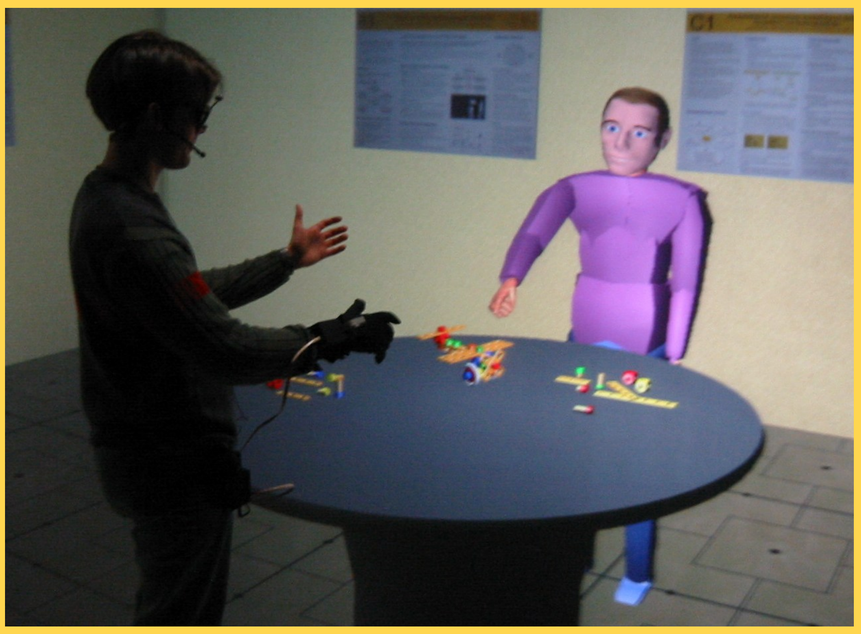
Resolution of Multimodal Object References using Conceptual Short-Term Memory

Thies Pfeiffer, Ian Voss,
SFB 360, {tpfeiffe; ivoss}@techfak.uni-bielefeld.de
Marc Latoschik,
A.I. & VR Lab, marcl@techfak.uni-bielefeld.de
University of Bielefeld, 33594 Bielefeld, Germany

Introduction

In the Collaborative Research Center (SFB 360) at the University of Bielefeld we are concerned wih situated artificial communicators. In one application scenario the user is involved in a task-oriented discourse with the embodied conversational agent MAX (see figure on the right). He guides him through an assembly process by means of task descriptions uttered using both German natural language and gestures. The work presented is on a system for identifying referent objects for deictic references in a real-time Virtual Reality (VR) Environment.



Application Scenario

- assembly task using toy kit
- the user is the instructor
- MAX is the constructor
- both can use multimodal utterances
- user and MAX are part of one environment
  - both can be used as referent object (e.g., "*the block on your right*")
  - or as frame of reference (e.g., "*the block to the left from your's point of view*")
- MAX supports the progress of the system
  - by pointing at the referenced objects
  - initiating a subdialogue for clarification

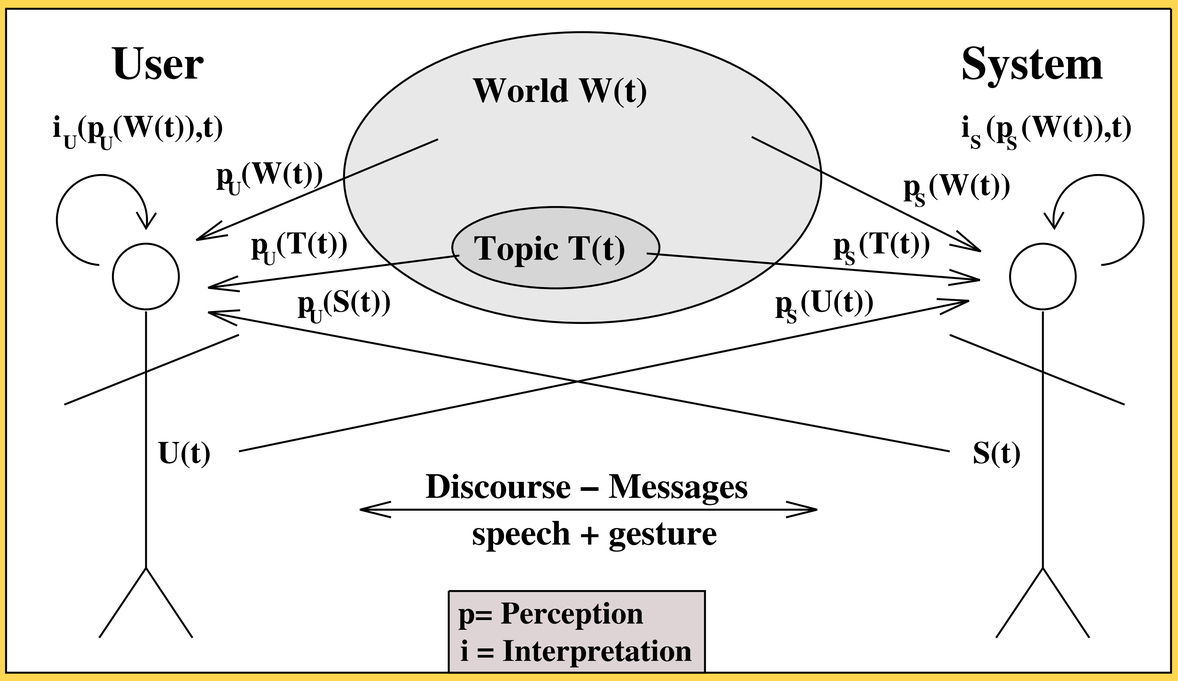
Reference Use

- instructions
  - are multimodal discourse messages
  - refer to a topic T in the world W
  - are perceived by the interlocutor (MAX)
- deictic references
  - refer to visualy perceptive objects in the world
  - are relative to a referent object and a frame of reference
- for reference resolution the propositional values of utterances are needed with respect to
  - the time of the specific utterance
  - the given relative scene representations
  - the view of the participants

Problems

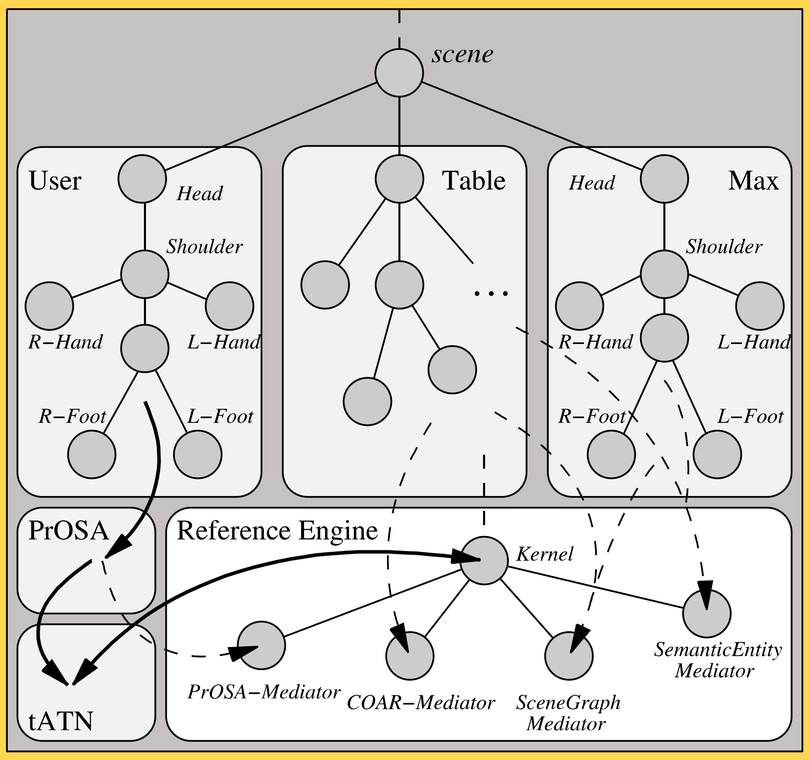
- positions and orientations (user, MAX, objects) continuously shift
- multimodal input is processed in different asynchronous channels
- the interpretation of percepts is always - at least mildly - running behind

It is neither plausible nor feasible to hold the complete scene information for each time frame over the complete cycle of uttering and understanding an instruction.



The figure shows a refined version of a model for reference use proposed in [1]. The user U produces an instruction using multimodal discourse messages referencing a topic T. The system S has to use perception and knowledge based methods to recognize the message and identify the topic.

A Reference Resolution Engine



The reference resolution engine (RRE) is embedded in a scene-graph, the basic modelling method for Virtual Reality applications. The ProSA engine and the tATN are analysing and parsing the utterances of the user. During this process the RRE identifies the referenced objects, supported by conceptual short-term memories which are part of the PrOSA engine.

We developed a framework for a reference resolution engine [2] based on:

- basically an extended Constraint Satisfaction Problem Solver
  - fuzzy evaluations for relative and ambigue referential expressions
  - hierarchical organization of constraints
  - support for explicit and implicit constraints (e.g., user preferences)
- findings from basic research, e.g., concerning
  - naming of objects and aggregates (proper names, types, functions)
  - spatial categorization [3]
  - alternative perspectives
  - recency effects
  - implicit user preferences and expectations

Conceptual Short-Term Memories

To cope with the problem of delayed evaluation, we introduced short-term conceptual memories, as proposed by [4], these are characterized as follows:

- several of such memories are bound to specific modalities
- keeping track of concepts for relevant objects and properties
- include relative positions
- the duration is as long as the time needed to process the instruction
- the capacity allows to hold as much undifferenciated concepts as possible
  - until they are selected by additional constraints
  - or filtered by a gradual "forgetting"

Conclusion

We use short-term conceptual memories for accessing and integrating perceptual information that is specific to different modalities. This significantly reduces the resource-costs of reference resolution and allows for real-time processing.

Acknowledgements

This work is part of the Collaborative Research Center (SFB360) at the University of Bielefeld and partly supported by the Deutsche Forschungsgemeinschaft (DFG).

References

[1] A. Popescu-Belis, I. Robba and G. Sabah, "*Reference Resolution beyond Coreference: a Conceptual Frame and its Application*", in Coling-ACL'98, Montreal, Canada 1998

[2] T. Pfeiffer, "*Eine Referenzauflösung für die dynamische Anwendung in Konstruktions-situationen in der Virtuellen Realität*", Diplomarbeit an der Technischen Fakultät der Universität Bielefeld, 2003

[3] C. Vorwerg, "*Raumrelationen in Wahrnehmung und Sprache*", Dt. Univ.-Verl, 2001

[4] M.C. Potter, "*Very short term conceptual memory*", Memory & Cognition, 21, 1993