Resolution of Multimodal Object References using Conceptual Short Term Memory

Thies Pfeiffer (tpfeiffe@techfak.uni-bielefeld.de)
Ian Voss (ivoss@techfak.uni-bielefeld.de)
Marc E. Latoschik (marcl@techfak.uni-bielefeld.de)
SFB 360; University of Bielefeld; 33594 Bielefeld, Germany

Introduction

This poster presents cognitive-motivated aspects of a technical system for the resolution of references to objects within an assembly-task domain. The research is integrated in the Collaborative Research Center SFB 360 which is concerned with situated artificial communicators. One application scenario consists of a task-oriented discourse between an instructor and a constructor who collaboratively build aggregates from a wooden toy kit ("Baufix"), or from generic CAD parts.

In our current setting this scenario is embedded in a virtual reality (VR) installation, where the human user, taking the role of the instructor, guides the artificial constructor (embodied by the ECA Max) through the assembly process by means of multimodal task descriptions (see Figure 1). The system handles instructions like: *"Plug the left red screw from above in the middle hole of the wing and turn it this way.*" accompanied by coverbal deictic and mimetic gestures (see Latoschik, 2001).

Problem of Reference Resolution in VR

While the relevant research mainly focusses both on static $2\frac{1}{2}$D or 3D scenes and references to single objects, our setting requests for a robust interpretation of full sentences with several possible references in an interactive and highly dynamic scenario. In VR, as in real life, egocentric or object intrinsic frames of reference might shift continuously during utterances due to movements of the user, of Max or of the objects. In consequence, references may only have a small duration of validity, which does not necessarily extend up to the end of the sentence.

A Reference Resolution System

We developed a framework (Pfeiffer, 2003), which allows us to incorporate findings from experiments in static settings concerning e.g. spatial categorization (Vorwerg, 2001), recency and several effects of user preferences.

The problem of the validity of references is stressed even more in technical systems, with a rather large timespan between the utterance of the user and the inference processing of the user instruction after the utterance has been fully recognized. To cope with this problem, we introduced short-term conceptual memories, as proposed by Potter (1993), into the system, which are bound to specific modalities. These memories incrementally keep track of concepts of objects of interest and their relative positions (what we call spacemaps) over a certain period of time. By utilizing this memories the inference system is capable not only to resolve the references due the time of processing, but also exactly in their time of validity - a crucial point for a correct understanding.

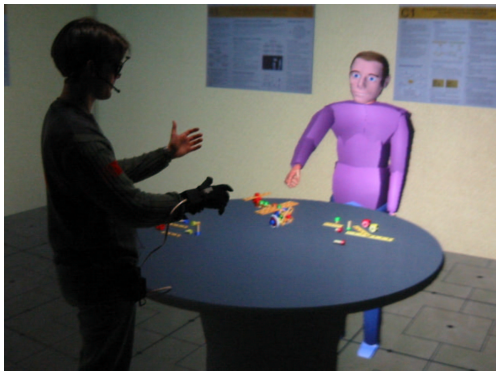

Figure 1: The user utters an object reference to Max.

Two questions about capacity and duration of those memories arose. At the moment we can only give rather pragmatic answers: The timespan is necessarily at least as long as the time needed to process the task description. And the capacity is tightly coupled with resources and processing capabilities. It should allow to hold as much undifferenciated concepts as possible, until they are, probably incremental, selected by additional constraints, e.g. induced through the user instruction.

References

Latoschik, M.E. (2001). *Multimodale Interaktion in Virtueller Realitaet am Beispiel der virtuellen Konstruktion*. DISKI volume 251. Berlin: infix.

Pfeiffer, T. (2003). *Eine Referenzaufloesung fuer die dynamische Anwendung in Konstruktionssituationen in der Virtuellen Realitaet*. Master thesis. Faculty of Technology, University of Bielefeld.

Potter, M.C. (1993). Very short term conceptual memory. *Memory & Cognition, 21*, pp. 156-161.

Vorwerg, C. (2001). *Raumrelationen in Wahrnehmung und Sprache*. Wiesbaden: Dt. Univ.-Verl.