Anticipation in a VR-based Anthropomorphic Construction Assistant

Ian Voss

University of Bielefeld
Artificial Intelligence Group
Faculty of Technology
D-33594 Bielefeld
ivoss@techfak.uni-bielefeld.de

Ipke Wachsmuth

University of Bielefeld
Artificial Intelligence Group
Faculty of Technology
D-33594 Bielefeld
ipke@techfak.uni-bielefeld.de

Abstract

We describe an implemented system that anticipates user instructions in a collaborative construction task in virtual reality. The functionality of the system is embodied by an anthropomorphic interface agent and enables the usage of functional names for pieces of an uncompleted target aggregate in natural language instructions. Based on an internal model of the construction state and salience ratings for aggregate parts, an anticipatory system behaviour is achieved that tolerates imprecise user instructions.

1 Introduction

In human-human communication one of the many facilitating features is that the interlocutor can anticipate what we want and thus enable efficient communication by bridging missing or underspecified information. We here describe an implemented system that anticipates user instructions in a virtual construction situation. It is a model-based anticipatory system according to the terms of [Ekdahl 2000]. The functionality of the system is embodied by the multimodal assembly expert MAX [Sowa, Kopp & Latoschik, 2000]. MAX acts as an anthropomorphic interface agent operating in large scale virtual reality. The agent is able to interact with the user in multimodal dialogue via speaking and gesturing to perform a collaborative construction task in a virtual environment. The anticipatory abilities together with MAX's abilities to handle discourse offer a system behaviour that is robust and allows imprecise user instructions to be compensated [Voss 2001]. The task performed is domain-specific and demonstrated using German instructions and the wooden toy kit "Baufix".

To give an idea Figure 1 on the left shows a partly-built aeroplane with the tail unit and the propeller still missing. There are three as yet unassembled parts in the construction scene, two three-hole bars and one bolt. An approach based on aggregate matching would allow recognition of the aggregate only when it is finished, here in Step 3 of Figure 1. Our anticipation model enables MAX to interpret the unassembled parts as parts of the target aggregate to be built, which allows the user to use the functional names of the parts from the beginning of a construction sequence.

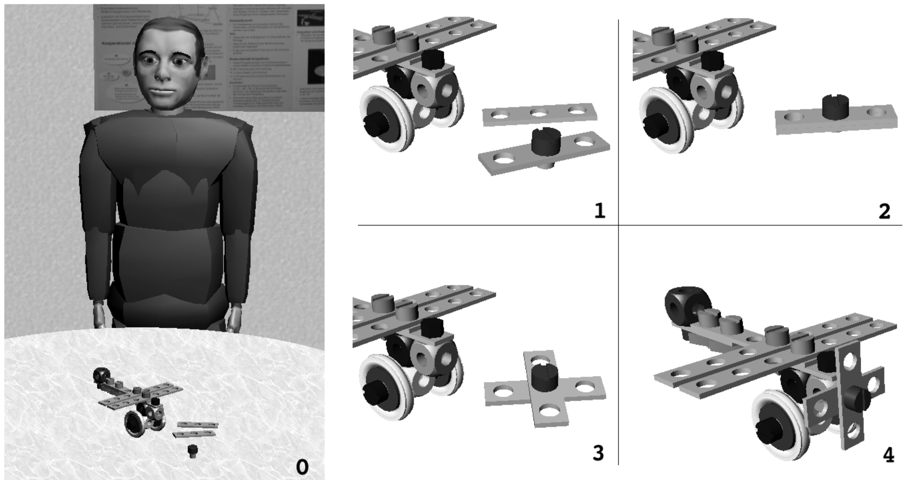


Figure 1: MAX in VR with partly-built aeroplane and unassembled parts on the left. On the right the progressive steps of the construction of a propeller are shown.

The interaction with MAX takes place in a fully immersive large scale virtual reality environment. The display technology used is a six channel three-sided high resolution stereo projection. Graphic rendering is done by a synchronized six-channel cluster based on distributed open-GL technology. The user interacts with the system by way of natural speech as well as deictic gesture. Data necessary for gesture interpretation are gained with the help of a marker-based real-time optical motion detection system. Wireless microphone and a speaker-independent incremental speech recognition system allow for natural speech input, auditory system reaction is done with an eight channel audio system to represent the location of an acoustic event in 3D space [Latoschik 2001].

2 Dynamic Conceptualization

The central idea for our solution is to use situated interpretations of knowledge bases to realize anticipation. The situated interpretation is done with respect to the state of construction as well as recent instructions to the system.

The knowledge bases used are of two kinds. The building parts are described in a "Baufix world" knowledge base (see Fig.2 on the left). Information about the different parts of the toy kit and all possible modes of interconnection are modelled in this knowledge base, such as the information that a possible connection between a bolt and a three-hole-bar would mean putting the shaft of the bolt through one of the holes with the head of the bolt not disappearing through the wooden bar. The possible models that can be build with these parts are described in an "Aeroplane world" knowledge base. It includes descriptions of functional aggregates, e.g., the tail unit or the propeller of an aeroplane (see Fig. 2 on the right).

Both these knowledge bases are implemented in the form of a hierarchically organized frame-structure [Jung 1998]. These knowledge bases are utilized in the process of virtual construction. For an adequate technical assistance of the user, a "join" action of two parts in the virtual construction scene is accompanied by the conceptualization of the newly formed aggregate. The conceptualization of an aggregate is done in the context of the construction situation of the toy aeroplane. Thus the actual interpretation of e.g. a wooden bar with three holes in it would be a propeller blade of an aeroplane when it is part of the completed and conceptualized assembly PROPELLER. (In another situation such a piece can be part of e.g. the tail unit.) A necessary prerequisite for a successful conceptualization is the completeness of an aggregate.

3 Anticipation

The frame-based knowledge representations mentioned above were originally designed to conceptualize, and thus to recognize, complete aggregates. This is necessary to technically follow the progress in construction. We now explain how a technical look-ahead in the construction situation is realized, by extending interpreting mechanisms to account for incomplete aggregates.

The "traditional" interpretation of the mentioned knowledge bases allows recognition of complete aggregates, i.e., only when each slot of a knowledge base concept is filled. Our new method enables us to continue using a frame-based representation even with some slots unfilled. When calculating anticipation, the role types are searched for the possible uses of a specific object type. Then this information is propagated back to the object type and where it is accumulated and standardised to generate salience ratings. This is done for all possible roles an object type can play.

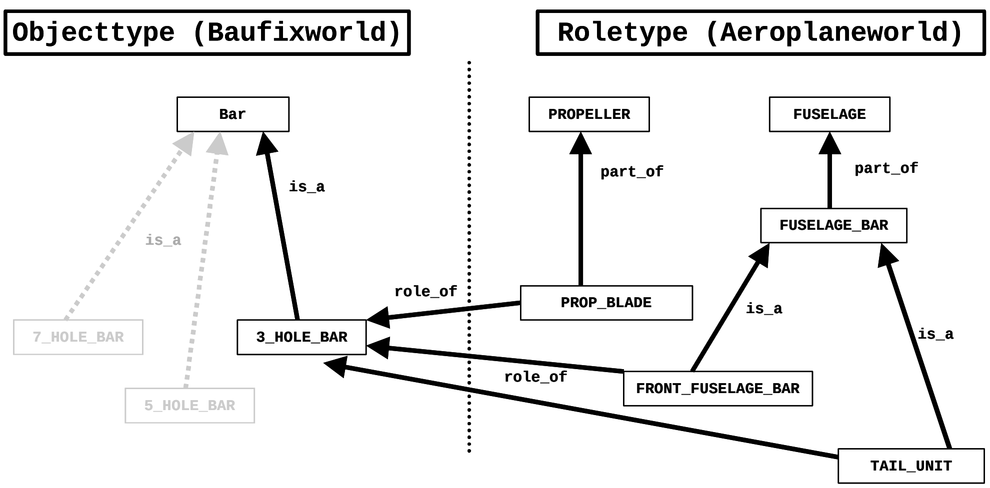


Figure 2: Knowledge bases of objects and roles

Anticipation is realized as a prospective interpretation of the possible places where to put a certain piece in the model. The interpretation of a single part is always done with respect to other as yet unassembled parts as well as to those parts already assembled to form an aggregate. Baufix parts receive salience ratings reflecting the likelihood of using this part in a certain role in one or another model. These ratings are dynamically calculated in the process of construction. Table 1 shows the ratings of the different parts and aggregates for the construction sequence of the propeller that is illustrated in Figure 1. Using this mechanism MAX is able to anticipate the possible roles of an object on the basis of the construction knowledge.

Table 1: Salience ratings for the anticipation of the parts and aggregates of the construction sequence in Figure 1

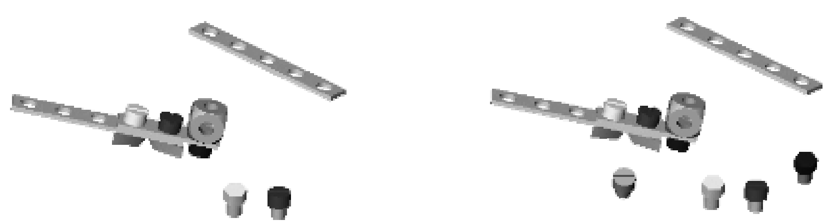
Step 0	Step 1	Step 2	Step 3	Step 4
PROP_BLADE 0,66	PROP_BLADE 0,66	PROP_BLADE 0,99	PROP_BLADE 1,00	PROP_BLADE 1,00
PROP_BLADE 0,66	PROP_BLADE 0,87	PROP_BLADE 0,99	PROP_BLADE 1,00	PROP_BLADE 1,00
PROP_BOLT 0,66 FUSELAGE_BOLT 0,66 TAILUNIT_BOLT 0,66	PROP_BOLT 0,87 FUSELAGE_BOLT 0,66 TAILUNIT_BOLT 0,66	PROP_BOLT 0,99 FUSELAGE_BOLT 0,66 TAILUNIT_BOLT 0,66	PROP_BOLT 1,00	PROP_BOLT 1,00
AEROPLANE 0,68	PROPELLER 0,66 FUSELAGE 0,20 AEROPLANE 0,68	PROPELLER 1,00 AEROPLANE 0,68	PROPELLER 1,00 AEROPLANE 0,68	AEROPLANE 0,79

Anticipation is done to enable the user to use the functional names of the parts of an aggregate that is still under construction when giving instructions to MAX. In the case of a propeller as the target aggregate, the user might as well use the functional name "propeller-blade" and not just "three-hole-bar" in an instruction like "put the bolt in the middle hole of the propeller-blade".

Anticipation is realized to understand functional names of aggregates or parts of them. This is implemented on the basis of salience ratings, which leaves the possibilities of two or more parts receiving the same ratings. In case of such ambiguities as well as simple underspecifications in user instructions, a dialogue system is used to clarify the user's intention.

4 Example

The anticipatory abilities of MAX represent an internal model of the state of construction. The full advantages of the robustness and situatedness in the construction situation can only be made use of when getting feedback for user instructions. The ability of MAX to ask context-sensitive questions is used to resolve ambiguities. Figure 3 shows two different behaviours of MAX in response to a user instruction like *"Put the tail-unit bolt in the middle hole of the five-hole bar"*. This instruction is even then underspecified when using anticipation because of identical salience ratings that do not enable MAX to choose the correct bolt by itself.



Question I: ***"The black or the white bolt?"*** Question II: ***"Which bolt do you want?"***

Figure 3: Context-sensitive question to the user

A user reaction to these questions shown in Figure 3 could be an underspecified instruction like *"The black one!"* for Question I or *"The left one!"* for Question II. The ellipsis in the answers is correctly interpreted with the use of a discourse memory.

5 Conclusion and Perspective

We here described a method to establish a mapping between an uncompleted aggregate, or even single parts of it, and the representation of the complete aggregate as described in the knowledge base. A corpus of human-human Baufix construction dialogues is available and exploited with respect to the usage of functional names or role types instead of object types. As anticipation is based on knowledge of role-type worlds, our approach is so far limited as to that no anticipatory assistance can be given for the construction of variants. The user is then still assisted in "join" actions when making use of the names of the object types. Discourse abilities that help disambiguating with or without the use of pointing gestures are not limited to the aeroplane world. The transfer to another domain is possible. Different role types have been modeled to show this such as a toy scooter. The transfer to another toy kit is limited to those toy kits making use of similar connection types between the building parts.

Follow-on work is underway to make construction episodes available for the next level of user instructions. It is intended to cope with instructions like "and now do the same thing on the other side" when just having built something that is needed in a similar way at another place on the aeroplane.

6 Acknowledgements

This work is part of the Collaborative Research Centre SFB 360 at the University of Bielefeld and partly supported by the Deutsche Forschungsgemeinschaft (DFG). Implementation assistance by Thies Pfeiffer is gratefully acknowledged.

7 References

Ekdahl, B. (2000), Anticipatory Systems and Reductionism. *Fifteen European Meeting on Cybernetic and Systems Research*, EMCSR 2000, Vienna, April 25-28.

Jung, B. (1998), Reasoning about Objects, Assemblies, and Roles in On-Going Assembly Tasks. *Distributed Autonomous Robotic Systems 3*, Springer-Verlag, 257-266.

Sowa, T., Kopp, S. & Latoschik, M. E. (2001), A Communicative Mediator in a Virtual Environment: Processing of Multimodal Input and Output. *Proc. of the International Workshop on Information Presentation and Natural Multimodal Dialogue*, pp. 71-74, Verona, Italy.

Latoschik, M.E. (2001), A gesture processing framework for multimodal interaction in virtual reality. In A. Chalmers and V. Lalioti, eds, "*1st Int. Conf. on Computer Graphics, Virtual Reality and Visualization in Africa,* pp. 95-100". ACM_SIGGRAPH.

Voss, I. (2001), Anticipation in construction dialogues. In J. Vanderdonckt, A. Blandford and A. Derycke, eds., "*Interaction without frontiers*", Proceedings of Joint AFIHM-BCS Conf. on Human-Computer Interaction. IHM_HCI 2001, Lille, France, volume 2 189-190.