

A Spoken Language Interface to a Virtual Reality System (Video)

Stephanie S. Everett, Kenneth Wauchope

Navy Ctr. for Applied Research in AI
Naval Research Laboratory
Washington, DC 20375, USA
everett|wauchope@aic.nrl.navy.mil

Manuel A. Pérez

Dept. Electrical & Computer Engineering
Recinto Universitario de Mayaguez, UPR
Mayaguez, PR 00681
mperez@exodo.upr.clu.edu

1 Description of video tape

Format: VHS. Duration: 9 min. 15 sec.

Immersive, interactive 3D computer display systems (often called virtual reality systems, or virtual environments) are rapidly emerging as practical options for training, command and control (C2), hazardous operations, visualization and other applications. However, the need for improved control and navigation techniques is well recognized (Herndon et al., 1994). The Navy Center for Applied Research in Artificial Intelligence has developed NAUTILUS (Navy AUTomated Intelligent Language Understanding System), a general-purpose natural language processing system, which has previously been integrated with the graphical user interface of a simulation-based C2 system to illustrate the advantages to be gained by combining natural language understanding (NLU) and direct manipulation in a human-computer interface (Wauchope, 1994). Using the NAUTILUS system in the interface to a virtual environment (VE) is a natural extension of this work.

The purpose of the project documented in this video was to demonstrate and explore some of the capabilities of a NLU interface to a VE system, and to identify some of the research issues that need to be addressed in this area. It is important to recognize that NLU is not simply speech recognition, where each individual utterance maps to a specific command. In a NLU system, a given sentence may have different meanings depending on the context, so a logical analysis of the utterance is required to determine the appropriate interpretation. This allows us to take advantage of certain powerful linguistic properties as described below.

One major difficulty with interfaces to VE systems is that the user's hands and eyes are occupied in the virtual world, so standard input devices such as mice and keyboards that require a physical support and/or visual attention are impractical. Joysticks,

gloves, and other manual input devices are useful for some types of control (pointing, manipulating objects), but they are not well suited to more abstract input functions. Language, however, is ideally suited to abstract manipulations; it is also the most natural form of communication for humans, and does not require the use of one's hands or eyes. It is especially useful for controlling things that do not have a physical presence in the VE, such as object scale, display characteristics, and time. It also provides a powerful means to access the knowledge that underlies the VE by allowing the user to ask questions of the system. Using speech output in combination with speech recognition helps to avoid the use of textual displays which can be difficult to read on immersive presentation equipment, and which can interfere with the user's view and the "reality" of the virtual world.

The prototype system shown in this film uses off-the-shelf speech recognition and synthesis technology combined with the NAUTILUS system and VIEWER (Solan and Hill, 1993), a 3D tactical scenario playback system developed by NRL's Tactical Electronic Warfare Division for a separate project. Building the prototype involved the creation of an application-specific dictionary and lexical semantics for NAUTILUS, a few minor extensions to its English grammar, and the development of two sets of code: one to translate the logical forms generated by NAUTILUS into messages for the application software, and one to interpret these messages and instruct VIEWER to produce the appropriate actions or responses.

The interface supports two classes of spoken input: commands and questions. The commands allow the user to control the playback of the simulation and its speed, as well as various display characteristics, such as viewpoint (*Show me the top-down/out-the-window view*) and overlays (*Display the map rings*). The user can also tell the system to hide or display

individual objects (*Show the Thunderbird*) or sets of objects (*Hide all the friendly aircraft that don't have missiles*). Commands are also used to manipulate time (*Run the simulation forward/backward; Set the clock to zero*). In addition, the user can move from one object to another by name or by description, rather than by flying or pointing (*Put me on the Doomsday; Put me on the hostile ship*), or by specifying a particular location (*Move me to 23 N, 40 E; Increase my altitude to 4000 feet*).

Questions allow the user to access information contained in the knowledge base that underlies the simulation. He/she can ask for information about the virtual world (*How many hostile ships are there?*), or about a specific object (*What is the Thunderbird's heading?*; *What is my viewing altitude?*) or set of objects (*Do any of the friendly ships have emitters on board?*). The user can also ask about the state of various aspects of the simulation (*Is the simulation running?*, *What is the time increment?*).

NAUTILUS keeps a history of prior references and their denotations, which allows the use of anaphoric reference (pronouns like *it* or *them*). It also supports relative clauses (*All the ships that have missiles on board*), and elliptical follow-ups involving substitution noun phrases (*How about the Titanic?*).

User feedback is handled in one of two ways: changes in the graphical display in response to commands, and verbal answers output through the speech synthesizer in response to questions. In the film, the answers consist primarily of yes/no and short noun phrases. The addition of a natural language generation module to generate appropriate verbal responses would improve feedback and smooth the flow of the discourse.

The grammar for the speech recognition module consists of a vocabulary of just over 300 words and a set of about 140 rules that support the recognition of approximately 1.2 million utterances. In order to make the interface flexible and easy to use, the rules have been designed to allow the user to phrase an utterance in a variety of ways, improving the naturalness of the interaction by taking advantage of the strong linguistic processing capabilities of NAUTILUS.

The use of natural language understanding in a human-computer interface can help provide the richness of control required in complex interactive environments. NLU provides a way to control abstract features such as time without the need for manual controls, and enables the user to access information in the underlying knowledge base(s) by asking questions. It also allows simulation objects to be referred

to by description, as sets, and generically rather than just individually. This film shows a demonstration of a prototype system that illustrates some of the capabilities of NLU in an interface to a virtual environment system. However, it must be stressed that this area of HCI is still in its infancy, and there are a number of research issues that will need to be addressed in order to realize the full potential of this technology.

References

- Kenneth P. Herndon, Andries van Dam and Michael Gleicher. 1994. The challenges of 3D interaction. *ACM SIGCHI Bulletin*, 26:4, pages 36-43.
- Brian T. Solan and Tobin A. Hill 1993. An application of object-oriented problem solving in electronic warfare simulation. NRL technical report NRL/FR/5707-93-9551.
- Kenneth Wauchope. 1994. Eucalyptus: integrating natural language input with a graphical user interface. NRL technical report NRL/FR/5510-94-9711.