Exploring Multimodal 3D Interfaces

S. Augustine Su *, Robert King *, and J. Edward Swan II

Naval Research Laboratory, Code 5580
4555 Overlook Avenue, SW
Washington, D.C. 20375-5337
{su, king, swan}@ait.nrl.navy.mil
* also affiliated with ITT Industries.

The Virtual Reality Lab at the Naval Research Laboratory has created a real-time, situational awareness virtual environment for battlefield visualization, termed the Dragon System [1]. The system is implemented on a workbench [2] and a CAVE-like immersive room [3]. It utilizes a *virtual laser pointer* metaphor for navigating and interacting with the terrain and objects; we implemented this metaphor with a hand-held joystick tracked with a 6-DOF magnetic tracker instrumented with several buttons [1]. Currently, we are extending this single-modal paradigm into various multimodal 3D interfaces, especially combining voice and gesture. Our objectives are exploring new multimodal interaction paradigms and understanding how users interact in virtual environments.

Recently, we have integrated the Oregon Graduate Institute's QuickSet [4] multimodal system with the Dragon System [5]. This work demonstrates a paradigm of using voice commands and ink gestures to accomplish tasks in 3D environments. For example, the user can utter "barbed wire" while drawing a line on either a 2D or 3D map, to create a "barbed wire" object in the 3D environment. We are currently exploring a related paradigm combining voice and the virtual laser pointer metaphor. In this paradigm, the user can say "put that tank on top of that hill", accompanied by first pointing to a selectable tank object while the word "tank" is uttered, and then pointing near a hill object while the word "hill" is uttered. The system then moves the tank to the top of the selected hill.

A major limitation in the use of virtual reality for practical applications has been the lack of a strong set of multimodal interface mechanisms. The most critical research issue is to demonstrate instances where multimodal interfaces have better performance and user satisfaction than traditional interface techniques. This line of research will inevitably yield a diversity of multimodal paradigms and techniques. Once we are so equipped, the next level of work is building a scalable architecture for supporting multimodal interactions. Different combinations of modalities may need different architectures. A good architecture should support the concept of dialog built upon tasks. One research issue is the temporal synchronization / coordination of modalities in a single interaction paradigm. For example, what time is the best moment for associating a deictic gesture with a key word or phrase in a voice command? Among different timing options, which are good for fast machine recognition? Which best support user comprehension? Another important issue, which is mostly ignored, is the granularity of modalities. For example, to consider voice commands, do we want to wait for the completion of a sentence before processing the information? Or, do we want to greedily utilize every bit of information that is possibly recognizable and available? Do we want to treat a hand gesture as a path of a single 6-DOF point? Or, do we want to exploit other knowledge about the hand, such as hand shapes for signifying information, or hand surface polygons for contacting objects? How do we coordinate this additional information? We believe these research issues are interesting and challenging.

[1]. Jim Durbin, J. Edward Swan II, Brad Colbert, John Crowe, Rob King, Tony King, Chris Scannell, Zachary Wartell, Terry Welsh, "Battlefield Visualization on the Responsive Workbench", *Proceedings of IEEE Visualization '98*, IEEE Computer Society Press, Oct. 1998, pp. 463–466.

[2]. Obeysekare, U., Williams, C., Durbin, J., Rosenblum, L., Rosenberg, R., Grinstein, F., Ramamurthi, R., Landsberg, A., and Sandberg, W., "Virtual Workbench --- A Non-Immersive Virtual Environment for Visualizing and Interacting with 3D Objects for Scientific Visualization", *IEEE Visualization '96 Annual Conference Proceedings*, IEEE Computer Society Press, Oct. 1996, pp. 345–349.

[3]. Carolina Cruz-Neira and Daniel J. Sandin and Thomas A. DeFanti. “Surround-screen Projection-based Virtual Reality: The Design and Implementation of the CAVE”, *Proceedings of SIGGRAPH 93, Computer Graphics Proceedings*, Annual Conference Series, Aug. 1993, pp. 135-142.

[4]. Philip Cohen, Michael Johnson, David McGee, Sharon Oviatt, Jay Pittman, Ira Smith, Liang Chen and Josh Clow, “QuickSet: Multimodal Interaction for Distributed Applications”, *Proceedings of the Fifth Annual International Multimedia Conference (Multimedia ’97)*, ACM Press, Nov. 1997, pp. 31-44.

[5]. Philip Cohen, David McGee, Sharon Oviatt, Lizhong Wu, Joshua Clow, Robert King, Simon Julier, and Lawrence Rosenblum, “Multimodal Interaction for 2D and 3D Environments”, *IEEE Computer Graphics and Applications*, Vol. 19, No. 4, July/August 1999, pp. 10-13.