

# MSVT: A VIRTUAL REALITY-BASED MULTIMODAL SCIENTIFIC VISUALIZATION TOOL

JOSEPH J. LAVIOLA JR.

Brown University Site of the NSF Science and Technology Center  
for Computer Graphics and Scientific Visualization  
PO Box 1910, Providence, RI 02912 USA

## ABSTRACT

Recent approaches to providing users with more natural methods of interacting with virtual environment applications have shown that more than one mode of input can be both beneficial and intuitive as a communication medium between humans and computer applications. Although there are many different modes that could be used in these applications, hand gestures and speech appear to be two of the most logical since users will typically be in environments that will have them immersed in a virtual world with limited access to traditional input devices such as the keyboard or mouse. In this paper, we describe a prototype application, MSVT (Multimodal Scientific Visualization Tool), for visualizing fluid flow around a dataset. MSVT uses a multimodal interface which combines whole-hand and voice input to allow users to visualize and interact with the dataset in a natural manner. A discussion of the various interaction techniques, and the results of an informal user evaluation are presented.

**KEYWORDS:** Multimodal Interaction, Virtual Environments, Scientific Visualization, Speech Recognition

## INTRODUCTION

Multimodal interaction provides many benefits over traditional unimodal metaphors such as WIMP (Windows, Icons, Menus, Point and Click) interfaces[1]. By combining whole-hand and speech input, human-computer interaction is augmented in a number of ways. Users can interact more naturally since human-to-human interaction often occurs with combinations of speech and hand movement. In addition, an application can achieve a better understanding of the user's intended action by providing it with multiple input streams because speech and whole-hand input cannot provide perfect recognition accuracy.

Combining whole-hand and speech input also has the

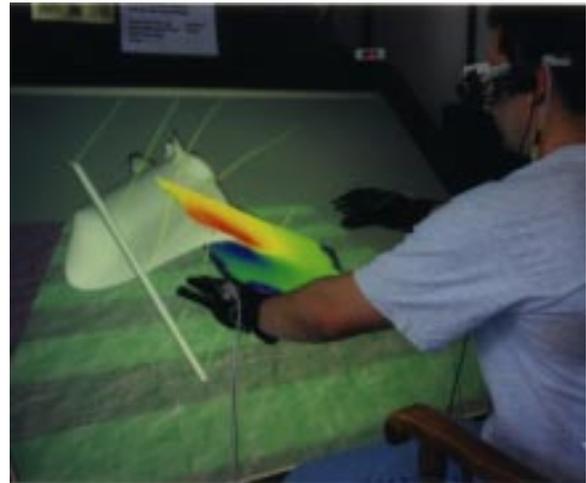


Figure 1: A user interacting with a dataset for visualizing a flow field around a space shuttle. The user simultaneously manipulates the streamlines with his left hand and the shuttle with his right hand while viewing the data in stereo.

advantage of simplifying the interface not only from the user's perspective but also from the developer's perspective. From the user's perspective, the interface can be simpler since one modality does not have to account for all interactions. For example, if user have to interact solely with speech or whole-hand input, they have to remember either a complicated speech vocabulary or a complicated gesture vocabulary. However, if we combine the modes in a complementary fashion, the set of interactions remains the same as either single modality, yet their respective vocabularies are simplified, easing cognitive load. By combining these two modalities we can also reduce recognition times and increase interaction speed since each individual recognition system has less work to do and takes less time in making decisions.

From the developer's perspective, the interface is somewhat simpler to implement in terms of algorithmic complexity. In order to provide a robust interface with either speech or whole-hand input (especially hand gestures), the developer would have to implement rather complex recog-

nitiation routines that would require many optimizations to provide fast interaction. Combining these two modalities splits the work allowing for a simpler implementation of each modal component.

Based on the advantages that multimodal interaction provides over traditional unimodal interfaces, we have developed a prototype application for visualizing and interacting with flow about a dataset (see Figure 1). Our Multimodal Scientific Visualization Tool (MSVT) is based on the premise that virtual reality provides an intuitive environment for exploring scientific data, an idea that was initially developed with Bryson's Virtual Windtunnel project[2, 3]. MSVT uses a rear-projected display device and combines speech input with pinching postures and gestures. The main objective of MSVT was not only to build a natural and intuitive interface for a scientific visualization application, but also to explore the following multimodal input combination styles[4]:

**Complementarity.** Two or more input modalities complement each other when they combine to issue a single command. For example, to instantiate a virtual object, a user makes a pointing gesture and then speaks. Speech and gesture complement each other since the gesture provides the information on where to place the object and the speech command provides the information on what type of object to place.

**Concurrency.** Two or more input modalities are concurrent when they issue different commands that overlap in time. For example, a user is navigating by gesture through a virtual environment and while doing so uses voice commands to ask questions about objects in the environment. Concurrency enables the user to issue commands in parallel; reflecting such real world tasks as talking on the phone while making dinner.

**Specialization.** A particular modality is specialized when it is always used for a specific task because it is more appropriate and/or natural for that task. For example, a user wants to create and place an object in a virtual environment. For this particular task, it makes sense to have a "pointing" gesture determine the object's location since the number of possible voice commands for placing the object is too large and a voice command cannot achieve the specificity of the object placement task.

**Transfer.** Two input modalities transfer information when one receives information from another and uses this information to complete a given task. One of the best examples of transfer in multimodal interaction is the push-to-talk interface[5]: the speech modality receives information from a hand gesture telling it that speech should be activated.

## ORGANIZATION

The remainder of this paper is organized in the following manner. The next section describes previous work related to multimodal interfaces and MSVT followed by a discussion of the application functionality. Then we discuss the results of an informal user evaluation. Finally, the last two sections provide areas for future work and a conclusion.

## PREVIOUS WORK

In the context of whole-hand and speech input, the use of a multimodal interface that integrates the two modalities can be traced back to Bolt's "Put That There" system[6] developed in 1980. This system used pointing hand postures and voice commands to create, manipulate, and edit simple 2D primitives such as squares and circles using a large rear-projected screen. Bolt extended his earlier work in 1992 with a multimodal interface that used hand gestures along with speech for manipulating 3D objects[7]. Weimer and Ganapathy developed another system that incorporated speech and hand gestures to create B-spline based 3D models[8]. However, their system was menu driven and did not take advantage of whole hand input. Other multimodal work that uses both hand gestures and speech can be found in[9, 10, 11].

Although there has been a significant amount work done using virtual reality for scientific visualization applications[2, 12, 13, 14, 15], the interaction paradigms used in these applications have been unimodal in nature. The combination of the Virtual Director[16] VR interface and the CAVE5D[17] visualization system is one of the few virtual reality-based scientific visualization applications that uses multimodal input. However, their system combines voice and wand input which limits the naturalness of their interface since users will typically only interact with one hand. By combining voice and whole-hand input using Pinch Gloves instead of a wand, MSVT allows users to interact with two hands which has been shown to be beneficial for many interaction tasks[18].

Frolich et. al.[19] developed a system for exploring geo-scientific data in virtual environments that led to the development of a multimodal user interface called the cubic mouse. The cubic mouse interface has visual, haptic, and audible components. The range of user interface problems the cubic mouse can address in the fluid-flow visualization domain is unclear. The cubic mouse is a useful tool for some specific problems such as precisely moving three orthogonal slicing planes. However, other tasks, such as creating a known visualization tool at a 3D location, are arguably more naturally performed with a speech and gesture interface.

## APPLICATION FUNCTIONALITY AND INTERACTION

MSVT gives users the ability to create, modify, drop, pick up, and delete a small set of visualization tools (streamlines, rakes, and colorplanes) for exploring the flow field about a given dataset. They also have the ability to change their viewpoint in the virtual environment, manipulate the dataset, make pictures of visualizations, and record and playback animations. Using the Fakespace Pinch Gloves, only the thumb, index, and middle finger on each hand and a set of speech commands are required to perform all the interactions in the application. The speech input component uses a vocabulary of over 20 voice commands. The following subsections describe the components of MSVT's interface in more detail including tool creation and manipulation, recording and playback, navigation, and dataset manipulation.

### TOOL CREATION AND MANIPULATION

MSVT provides three visualization tools; streamlines, rakes (see Figure 2), and colorplanes for exploring the fluid flow around the dataset. In order to create a given tool, users simply extend their arm to the display device and ask for the appropriate tool as shown in Figure 3. The hand that has the greatest distance from the tracking device's transmitter is the hand the object attaches to. For example, to put a colorplane in the right hand, simply extend the right arm and say "COLORPLANE". The colorplane is then instantiated and attached to the right hand where it can be moved through and around the dataset. This type of interface presents a natural way to instantiate tools by using a "show and ask" metaphor which utilizes both complementarity and transfer multimodal input styles. Users simply ask the application for some object and show where the object is to go.

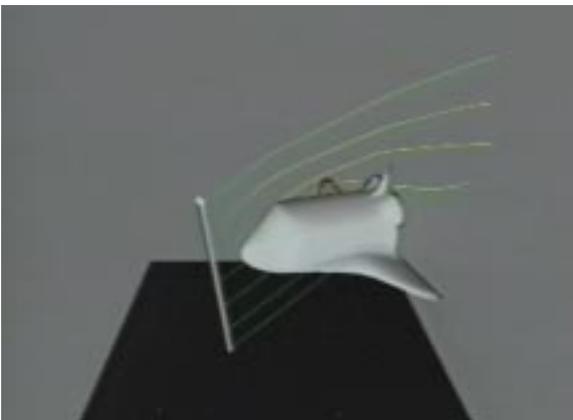


Figure 2: The rake visualization tool which is made up of a number of streamlines.

The visualization tools can be similarly dropped into



Figure 3: The user extends his right hand to the display asking for a streamline.

the environment: users move the tool to where it should be located while uttering a "DROP *object*" command where *object* is equal to streamline, colorplane or rake. Users can then manipulate the dataset with the visualization tool fixed. If users want to pick up a fixed object they simply hold out the appropriate hand and utter the "PICK UP" command asking for a particular tool. Visualization tools can also be deleted from a hand by holding out that hand and asking the application to remove that tool with the "REMOVE" command.

Once a rake or colorplane has been created, users can change the number of streamlines attached to the rake and the size of the colorplane. These parameter changes are made using a digital input slider, a slider which has no analog components. It simply consists of some type of button and the ability to determine the relative positions of itself and the entity that invokes the button press. The Pinch Gloves can make a very effective digital input slider since the conductive cloth patches attached to each fingertip extend down the back of the each finger as well. Therefore, the user can make a connection by sliding one fingertip along the back of another. Using the trackers attached to each hand we can then determine whether the user slides a finger along the back of another in a direction moving toward the wrist or away from the wrist. Knowing this direction, we can then increase or decrease a parameter value to one of the visualization tools. So, the number of streamlines attached to a rake can be increased or decreased with the left index finger/right index finger slider and the size of a colorplane can be increased or decreased with the left index finger/right middle finger slider. These increases and decreases are fixed in the application. Currently, two streamlines are added or removed from a rake depending on the direction of slider manipulation, and the colorplanes are increased or decreased in size by a factor of 0.5. Note that these values could be changed dynamically based on user preference if appropriate speech commands were in place.

A question arises as to how to determine what tool's parameter value to change if there is one in each hand; a rake in each hand, for example. This issue is resolved by holding the hand stationary that has the intended object of interest. With this approach, the hand that has moved the least during the course of the slider manipulation indicates which tool to modify. So, if a rake is in each hand and to increase the number of streamlines attached to the rake in the left hand by four, users can hold the left hand fixed and then slide the right index finger along the back of the left index finger, away from the wrist. Doing this twice would add four more streamlines to the rake in the left hand.

## RECORDING AND PLAYBACK

Having the ability to save and reinvestigate certain portions of a visualization session is an important part of using visualization techniques to better understand scientific data because it allows scientists to go back and reexamine interesting visualizations and show them to colleagues and collaborators. Therefore, MSVT provides two mechanisms for saving and retrieving visualizations. The first takes snapshots of a given scene by simply asking the application to "REMEMBER THIS VIEW". The view can then be retrieved with the "SHOW ME SAVED VIEW" command. The second mechanism records and plays back interaction animations. When the "START RECORDING" command is issued, the background color of the screen turns red, as shown in Figure 4, indicating that the application is in recording mode. Users then make the animation and say "STOP" to finish the recording. To view the animation users can issue the "PLAYBACK" command. During playback, the background color turns green, shown in Figure 5, indicating the mode change. These recording and playback tools not only benefit users who want to go back to previous visualizations, but also in collaborative settings when they need to show colleagues important visualizations they have discovered.



Figure 4: The user is in recording mode as indicated by the red background.

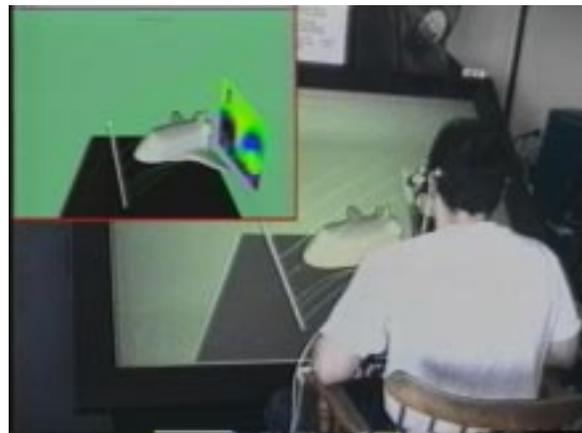


Figure 5: The user watching a previously recorded animation indicated by the green background.

In many cases, users will have both hands occupied interacting with the dataset but may still need to issue commands. For example, if speech were not available, when manipulating a tool in one hand and the dataset in the other, users would have to free one hand and make a gesture or press a button to start a recording. MSVT solves the problem by taking advantage of the concurrent multimodal input combination style of speaking and direct manipulation.

## NAVIGATION

Users navigate through the virtual environment with two possible interaction tools. Based on Multigen's SmartScene navigation techniques [20, 21], users can pull themselves through the virtual world by pinching the thumb and middle finger on either hand and grabbing a point in space. Translation is not constrained so movements in  $x$ ,  $y$ , and  $z$  can be made. When the users invoke the technique with one hand after the other, they can virtually walk through the VE.

Users can also pinch the thumb and middle finger of each hand simultaneously which results in the ability to scale, rotate, and translate the virtual world in one motion. This technique can be thought of as three distinct components (see Figure 6). First, scaling the viewing region by moving the two hands closer or farther apart along a fixed line. Second, rotating the world by making arc-like motions with each hand in opposite directions or keeping one hand stationary and making arc-like motions about the stationary hand. Third, translating about the virtual world by moving both hands simultaneously keeping the distance between the hands constant throughout the operation. By combining these three components users can perform scaling, rotation, and translation in one motion. For example, moving to a specific location and facing a certain direction while zooming to a close up view of an area of interest in the dataset.

Another method for navigating about the virtual envi-

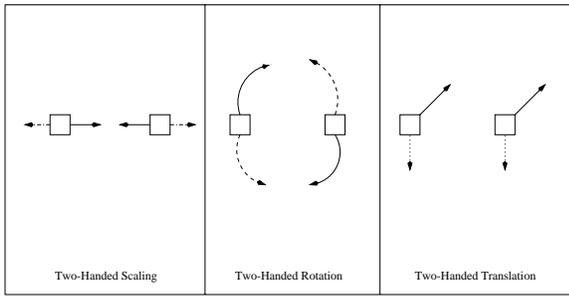


Figure 6: The three basic components of the two-handed navigation technique. The boxes represent the user’s hands and the line styles represent possible motions. These components can be used in isolation or by combining them so the viewing region can be scaled, rotate and translated in one motion. Note that using one hand at a time also allows for translation.

ronment is using the hand as a palette. In our case, the right hand acts as the palette while the left index finger is used to instantiate interactions. When users touch the left index finger to the right pinkie, middle finger, or thumb, the world rotates about the center of the dataset by 90, 180, and -90 degrees respectively. This navigation tool can be used for quickly rotating around the dataset.

## DATASET MANIPULATION

The task of dataset manipulation, specifically rotation and translation, is performed with the thumb and index finger on each hand. With the right hand, when users touch the thumb and index finger, they can translate and rotate the dataset by simply moving and rotating the hand. This manipulation provides six degrees of freedom. To perform rotation or translation of the dataset in isolation, and users touch the thumb and index finger of the left hand. Either dataset rotation or translation is performed based on the initial angle of the user’s hand about the  $z$  axis. If the hand angle is approximately zero degrees about the  $z$  axis, (i.e. the tracker attached to the back of the hand is approximately parallel with the floor) dataset rotation is performed otherwise the dataset is only translated. Note that touching the thumb to the index finger with both hands simultaneously is analogous to the two-handed navigation technique described in the previous subsection, except scaling is omitted.

## USER AND PROTOTYPE EVALUATION

Throughout the life of MSVT, a number of people, from academia, industry, and government, both have tried and observed the application. In general, users found the application to be compelling in terms of the interaction and the

virtual environment. From an interaction perspective, the majority of users found the interface easy to use (with some training) and liked the idea of the “show and ask” metaphor for creating and selecting visualization tools. The “show and ask” metaphor is an improvement over other object creation and selection techniques such as aperture-based selection[22] and 3D menus because these techniques require users to go to the object of choice in order to select it or to navigate through many layers of 3D menus to create an object. The “show and ask” metaphor is a faster method of creating and selecting these visualization tools since users do not have to actively select the virtual objects; with “show and ask” they come to the user via simple voice commands and are properly positioned by showing the application which hand to place the object in.

Users also found the digital sliders for increasing or decreasing the number streamlines attached to a rake and increasing or decreasing the size of a colorplane to be a simple yet effective way to manipulate these parameters. In addition, the majority of the users found the recording and playback capabilities to be an important part of the application. From the virtual environment perspective, most of the users (especially those with little or no VR experience) found the semi-immersive display to be extremely compelling from a visual standpoint. Users seemed to enjoy the stereo display and often tried to physically touch the virtual objects which is a good indicator that the stereo effect is working well.

Users provided a number of useful suggestions and constructive criticisms for improving the application. A number of users wanted other visualization tools in addition to the ones provided. Specifically, a number of people wanted the application to provide a form of text output that showed exact values in the flow field based on user input. Other tools that users thought might make the application more robust were particle traces and isosurfaces. Another important suggestion had to do with scaling with respect to interaction within the flow field. Users wanted to scale down the movement of the visualization tools so fine grained manipulation could be performed when users were close to the dataset.

Another important comment was that even though they found the application to be visually compelling, they were uncertain as to whether it would actually help them to be better scientists and to understand the data easier and more efficiently. This question not only plagues MSVT but all virtual environment-based scientific visualization systems. Unfortunately, the question is difficult to answer and is a definite area for future work and consideration.

Finally, one of the major problems with MSVT that people commented on had to do with the speech recognition. One of the goals of MSVT was to see how effective a natural speech interface without any push-to-talk mechanisms would function using current technological compo-

nents. In an isolated environment (an environment where only the user of the system is present), the speech recognition worked well and there were very few problems with false positive recognition. However, in a collaborative or demonstration scenario, the speech recognition often broke down due to environmental noise, recognizing words it wasn't supposed to causing erroneous operations. In some cases, these speech recognition problems made the application unusable. The main reason these problems occurred was that the application could not distinguish between the user speaking to it and to other people in the environment. As a result, with the current state of technology, we concluded that we could not have the user effectively interact with the application at a level of communication that mimics face-to-face human conversation. Therefore, an intermediary in the form of a push-to-talk interface is required when users are in collaborative or demonstration settings.

As a result, we added a voice command to the interface which triggered the speech recognition engine. The command "COMPUTER START LISTENING" was used to tell the application to listen to the user's voice commands while "COMPUTER STOP LISTENING" was used to tell the application to ignore all voice commands except for the voice activation command. This push-to-talk interface worked well, but other less intrusive mechanisms are required to make the application more usable.

## FUTURE WORK

Since MSVT is a prototype application, a significant amount of future work remains in order to make our system robust. Based on the comments from many of the users that have tried MSVT, we plan to introduce more visualization tools including text-based output so users can see values for specific quantities. We also plan to add scaling control mechanisms so the movement of visualization tools will be based on the user's size and proximity to the dataset. In addition, we plan to investigate the best way to incorporate a push-to-talk mechanism into the application which comes as close to face-to-face communication as possible.

Besides making additions to the application, an important area of future work is to determine the benefits of MSVT by conducting formal user evaluations. One of our goals in these formal evaluations is to determine if scientists have a better understanding of the information they are presented with using MSVT over other traditional desktop applications.

## CONCLUSIONS

In this paper, we have presented MSVT, a virtual reality-based multimodal scientific visualization tool for examin-

ing fluid flow about a dataset. By providing the user will a multimodal interface combining voice and two-handed input, we allow them to take advantage of communication skills that they have had a lifetime to acquire. In addition, users can interact with the application with both hands and still issue relevant commands using speech. We have also explored the use of multimodal combination styles such as complementarity, concurrency, specialization, and transfer and how they can be used in the context of a scientific visualization application. With further study and research, it is our goal to continue to find new ways to use multimodal interaction in virtual reality-based scientific visualization so that scientists and researchers can be more productive and efficient in their work.

## ACKNOWLEDGMENTS

Special thanks to Steve Bryson for providing the space shuttle dataset and to Timothy Rowley and Andries van Dam. This work is supported in part by the NSF Graphics and Visualization Center, IBM, Advanced Network and Services, Alias/Wavefront, Microsoft, Sun Microsystems and TACO.

## REFERENCES

- [1] A. van Dam. Post-WIMP User Interfaces. *Communications of the ACM*, 40(2), 1997, 63-67.
- [2] S. Bryson. Virtual Reality in Scientific Visualization. *Communications of the ACM*, 1996, 39(5):62-71.
- [3] S. Bryson, S. Johan, and L. Schlecht. An Extensible Interactive Visualization Framework for the Virtual Windtunnel. In *Proceedings of the Virtual Reality Annual International Symposium*, 1997, 106-113.
- [4] J. C. Martin. TYCOON: Theoretical Framework and Software Tools for Multimodal Interfaces. *Intelligence and Multimodality in Multimedia interfaces*. (ed.) John Lee, AAAI Press, 1998.
- [5] J. J. LaViola Jr. Whole-Hand and Speech Input In Virtual Environments. Master's Thesis, CS-99-15, Brown University, Department of Computer Science, Providence, RI, December 1999.
- [6] R. A. Bolt. Put That There: Voice and Gesture at the Graphics Interface. In *Proceedings of SIGGRAPH '80*, ACM Press, 1980, 262-270.
- [7] R. A. Bolt and E. Herranz. Two-Handed Gesture in Multi-Modal Natural Dialog. In *Proceedings of the Fifth Annual ACM Symposium on User Interface Software and Technology*, 7-14, 1992.

- [8] D. Weimer, and S. K. Ganapathy. Interaction Techniques Using Hand Tracking and Speech Recognition. In *Multimedia Interface Design*, Meera M. Blattner and Roger B. Dannenberg, (eds.), Addison-Wesley Publishing Company, New York, 1992, 109-126.
- [9] H. Ando, Y. Kitahara, and N. Hataoka. Evaluation of Multimodal Interface using Spoken Language and Pointing Gesture on Interior Design System. In *International Conference on Spoken Language Processing*, 1994, 567-570.
- [10] M. Billinghurst, J. Savage, P. Oppenheimer, and C. Edmond. The Expert Surgical Assistant: An Intelligent Virtual Environment with Multimodal Input. In *Proceedings of Medicine Meets Virtual Reality IV*, 1995, 590-607.
- [11] D. B. Koons, C. J. Sparrell, and K. R. Thorisson. Integrating Simultaneous Input from Speech, Gaze, and Hand Gestures. *Intelligent Multimedia Interfaces*, (ed.) Mark T. Maybury, 1993, 257-279.
- [12] G. H. Wheless, C. M. Lascara, A. Valle-Levinson, D. P. Brutzman, W. Sherman, W. L. Hibbard, and B.E. Paul. Virtual Chesapeake Bay: Interacting with a coupled physical-biological model. *IEEE Computer Graphics and Applications*, 1996, 16:52-57.
- [13] C. Cruz-Neira, J. Leigh, C. Barnes, S. M. Cohen, S. Das, R. Engelmann, R. Hudson, M. E. Papka, T. Roy, L. Siegel, C. Vasilakis, T. A. DeFanti, and D. J. Sandin. Scientists in Wonderland: A Report on Visualization Applications in the CAVE Virtual Reality Environment, In *Proceedings of IEEE 1993 Symposium on Research Frontiers in Virtual Reality*, 1993, 59-66.
- [14] D. Song and M. L. Norman. Cosmic Explorer: A Virtual Reality Environment for Exploring Cosmic Data. In *Proceedings of IEEE 1993 Symposium on Research Frontiers in Virtual Reality*, 1993, 75-79.
- [15] M. Jern and R. A. Earnshaw. Interactive Real-Time Visualization Systems Using a Virtual Reality Paradigm. *Visualization in Scientific Computing*, (eds.) M. Gobel, H. Muller, and B. Urban, Springer-Verlag, 1995, 174-189.
- [16] <http://virdir.ncsa.uiuc.edu/virdir/>
- [17] <http://www.ccpo.odu.edu/~cave5d/homepage.html>
- [18] W. Buxton and B. Myers. A study in two-handed input. In *Proceedings of CHI '86*, 1986, 321-326.
- [19] B. Frohlich, S. Barrass, B. Zehner, J. Plate, and M. Gobel. Exploring Geo-Scientific Data in Virtual Environments. In *IEEE Proceedings of Visualization '99*, 1999, 169-173.
- [20] D. J. Mapes, and M. J. Moshell. A Two-Handed Interface for Object Manipulation in Virtual Environments. In *PRESENCE: Teleoperators and Virtual Environments*, 1995, 4(4):403-416.
- [21] Multigen. SmartScene™ Video Clip, Discovery Channel's NextStep program, 1998.
- [22] A. S. Forsberg, K. Herndon, and R. C. Zeleznik. Aperture Based Selection for Immersive Virtual Environments. In *Proceedings of the 1996 ACM Symposium on User Interface Software and Technology*, 1996, 95-96.