

# SANDBOX : an Interface to Scientific Data based on Experimentation

*Andrew Johnson and Farshad Fotouhi*

Wayne State Univ., Department of Computer Science, Detroit, MI. 48202, USA  
(313) 577-3107 voice, (313) 577-6868 fax  
fotouhi@cs.wayne.edu

*Jason Leigh and Thomas DeFanti*

Univ. Illinois at Chicago, Electronic Visualization Laboratory, Chicago, IL. 60680, USA  
(312) 996-3002 voice, (312) 413-7585 fax  
spiff@eecs.uic.edu

KEYWORDS: Scientific Databases, Interfaces, Virtual Reality, Browsing

## ABSTRACT

Scientific databases contain very large amounts of data accessed by investigators from many disciplines. In this paper, we propose a new interface to scientific databases: the SANDBOX: Scientists Accessing Necessary Data Based On eXperimentation. Much of the data that is stored in scientific databases is collected through experimentation. The SANDBOX is a virtual reality tool which allows an investigator to 'recreate' the original experiment, collecting data from the scientific database in much the same way that the original data was collected. The investigator places virtual instruments into a virtual environment and collects data from the scientific database without ever typing in a query. These instruments give feedback, allowing the user to browse through available data of any type. We have implemented a prototype of the SANDBOX on a subset of NASA's FIFE scientific database using the CAVE virtual reality theatre.

## 1 Introduction

In recent years there has been much work devoted to collecting and analyzing large amounts of data from scientific experiments. This trend will continue into the future with even larger amounts of data being collected, stored, and analyzed. A principle problem facing users of such information systems is finding the appropriate information without detailed knowledge of the structure of the stored data.

Scientific databases are accessed by users from a wide range of disciplines, mostly unfamiliar with databases and their associated query languages. These users need to search for specific pieces of data quickly. They need to browse through related information to see if it is of value to them. They need to relate information from different parts of the database.

Much of the difficulty in accessing data in scientific databases comes from the enormous amount of data that is involved; but the organization of this data is also a major problem. Users from various scientific communities see different relationships between sets of data. Typically a generic interface is provided. This gives researchers from all backgrounds access to the data, but each of the researchers must conform to this generic structuring of the data.

Using virtual reality [3], visualization, and hypertext [11] we can hide the database from the user. Typically virtual reality and visualization are used to analyze data after it has been extracted from a scientific database, but they can also be used to make retrieving data from the database easier, and more intuitive. This virtual reality interface includes hypertext-like linkages between different types of data (numeric, graphical, and meta-data). We call this interface the SANDBOX: Scientists Accessing Necessary Data Based On eXperimentation.

An investigator can use virtual reality to ‘recreate’ the original experiment, collecting data from the scientific database in much the same way that the original data was collected. The investigator places virtual measuring devices into a virtual environment and collects data from the scientific database without ever typing in a query. Virtual reality provides more realistic interaction with the experiment than standard 2D displays, giving the user the feeling of being a participant in the experiment, not just looking at it from outside. As the user adds experiments to the sites, the instruments give the user feedback. This allows the user to browse through the data stored in the scientific database.

Scientific databases often contain additional information that can not be put into the rigid structure of typical databases: notes, drawings, diagrams, maps, photographs, sounds, etc. This data can be seamlessly included in the virtual reality interface, giving the user additional information of any kind. It is important to link together related information from multiple media sources to give each user access to all the available information. As the user recreates the experiments, the data is retrieved from the appropriate source, allowing the user to browse through all the data.

Section 2 discusses the problems with current scientific database interfaces. Section 3 discusses our proposed interface. Section 4 discusses our implementation of the SANDBOX paradigm. Section 5 discusses user reaction to this interface. Finally, Section 6 gives our conclusions and plans for future work.

## 2 Scientific Databases

Relational databases are designed to deal with limited ranges of data on specific topics. The form of the data is known ahead of time. The database tables and their relationships (the database schema) are clearly defined before the data is entered. Scientific databases [4, 10] contain a much larger amount of data on many different, but related topics. The form of the data is not known ahead of time as the data is collected by investigators from a wide range of disciplines. Scientific data is often stored using relational databases. This huge amount of interrelated data is forced into the rigid table structure of a relational database which can not adequately model the necessary relationships [7].

Unfortunately, each discipline stores data in its own way, making it very difficult for investigators in other fields to access it. Users from various scientific communities see different relationships between sets of data. Certain information is important to certain investigators and certain information is not.

When an interface to a scientific database is designed, its creator typically imposes a generic structure on the database - a hierarchical menu system allowing the user to move through an ordering of the tables. This gives researchers from all backgrounds a way to access the data, but each of the researchers must conform to this generic structuring of the data. This approach has several shortcomings: 1) The menu system does not provide enough flexibility for a wide range of researchers, 2) The users may not know enough about the domain to make appropriate choices, 3) It does not help the user with ill-defined queries.

Graphical query languages have been proposed to simplify the interface [8]. Graphical query languages make the database schema more visible, reduce typing, and allow users to rely on recognition rather than memorization. This approach has several shortcomings: 1)

The schema of scientific databases are so large, and complicated that the user rapidly runs out of screen real-estate, 2) The graphical metaphor quickly becomes cumbersome for complicated queries.

Ioannidis, et al [5] developed a graphical interface for the management of scientific experiments and data using the Object-Oriented data model MOOSE. The user interacts with the database through the schema. The system makes large schemas more manageable by allowing the user to hide parts of the schema, collapse sections of the schema into nodes, and use reference nodes to eliminate long arcs. While useful for scientists involved in the original experiment, this approach has several shortcomings for users less familiar with the original experiment: 1) The users may not know what data is available, 2) It gives users a variety of choices without sufficient descriptive material to make that choice, 3) The original schema may not match the relationships seen by all users.

Ahlberg, et al [1] experimented with using graphical widgets to formulate database queries. Graphical visualization was used to show the contents of the small database (the periodic table) and the results of the queries. By hiding the database schema and allowing the user to interact directly with the data values, they found that the users to gained a faster understanding of the data than with queries based on textual interfaces. Expanding on this, and allowing the user to have a more realistic interaction with the data values of a much larger database should give the user a more intuitive way of accessing their data.

Hypertext has been proposed as a way to give users the capability to browse through the meta-data associated with scientific databases [16]. This gives the users better understanding of the contents and organization of the database. Expanding on this, and allowing the user to browse through the data in the database as well as the meta-data should give the user a better understanding of the relationships among the various data.

### **3 SANDBOX**

Much of the data that is stored in scientific databases was collected through experimentation. Using virtual reality an investigator can ‘recreate’ the original experiment, collecting data from the scientific database in much the same way that the original data was collected. The investigator places virtual instruments into a virtual environment and collects data from the scientific database without ever typing in a query. Each investigator uses familiar measuring instruments and collects data in a familiar way. Using virtual reality, visualization, and hypertext we can hide the scientific database from the user.

This gives the user a ‘virtual laboratory’ that can be configured for different experiments on different scientific databases by loading in different sets of instruments, and environments. The exact instruments, and the way space and time are modeled, will depend on the individual experiment. The laboratory can become as large as the universe or small as an atom, it can move through time or space, depending on the experiments being run inside it. Space can be measured in angstroms, miles, or light years. Time can be measured in nanoseconds, days, or millennia. Since the instruments are virtual, they can be calibrated to display their values in whatever scale the user chooses.

As the user places instruments into the environment, they react as the real instruments would. The instruments give the user feedback. The investigator can use this feedback to add additional instruments to the experiment, move the instruments to other locations, or remove unnecessary instruments. These instruments allow the user to visualize the contents of the database before any actual data is retrieved, so the user can browse through the data. Once the user has placed the appropriate instruments into the environment and set the appropriate time interval, the information is retrieved from the database and stored in an external file for further use.

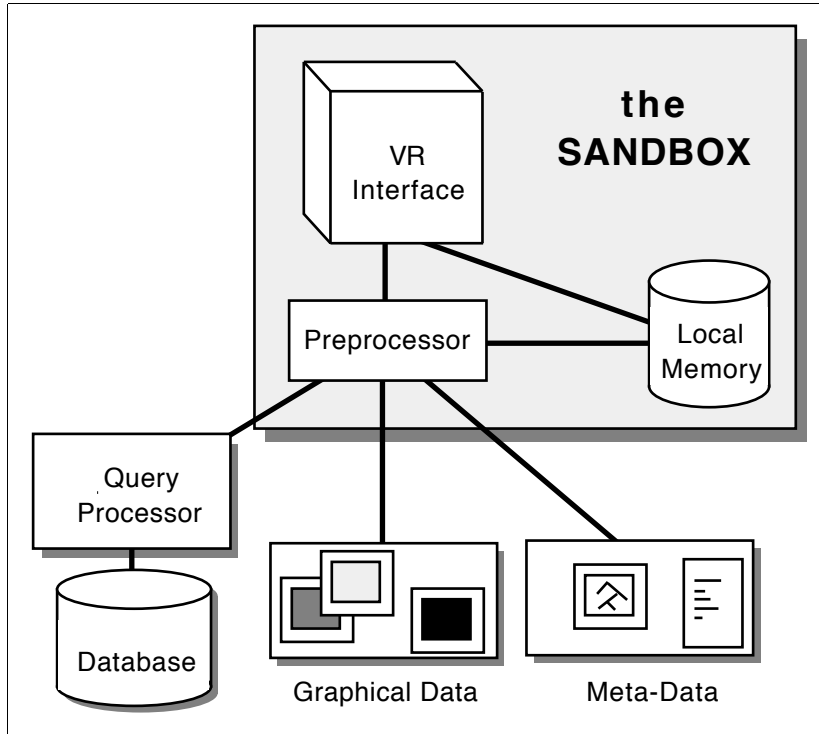


Figure 1: System Diagram of the SANDBOX

The interface employs an object-oriented paradigm. Existing virtual reality systems [15] allow for the creation of sliders, menus, and button panels. These are simply transferences of 2D systems into the 3D virtual environment where operations on objects are performed using a separate detached interface control panel. In our approach each object in the virtual environment maintains its own state, and has its own interface control panel. This reduces user errors during interaction by giving users access only to those functions which are appropriate to the object [12].

An overview of the SANDBOX is shown in Figure 1. The database management system must become more sophisticated to deal with the user's naivete. Using this virtual reality interface the user is given access to all the available data in the database. A single instrument may be linked to multiple attributes in multiple tables of the database.

Current database access methods are not fast enough to support the needs of virtual reality [13]. Virtual reality requires very fast access times. For smooth movement of a three dimensional image, at least 15 frames per second must be generated for each eye [20]. Generating a frame involves accessing all the relevant information, converting it into graphical form, and drawing it. Relevant portions of the database need to be brought into local (fast) memory before the visualization can begin, or as the visualization is proceeding in a form of progressive refinement. Typically, during visualization all of the necessary information is loaded into RAM before the visualization begins. This is clearly not possible here given the huge amount of data involved. Since visualization is used while retrieving information, not just afterwards, the entire database needs to be accessible.

In our system, as the user chooses instruments, times, and other experimental parameters, the preprocessor determines which parts of the database are likely to be accessed in the near future. Relational tables can be partitioned vertically and horizontally, objects can be isolated, and files can be marked. These blocks of information can then be moved into local memory before they are needed by the visualization system.

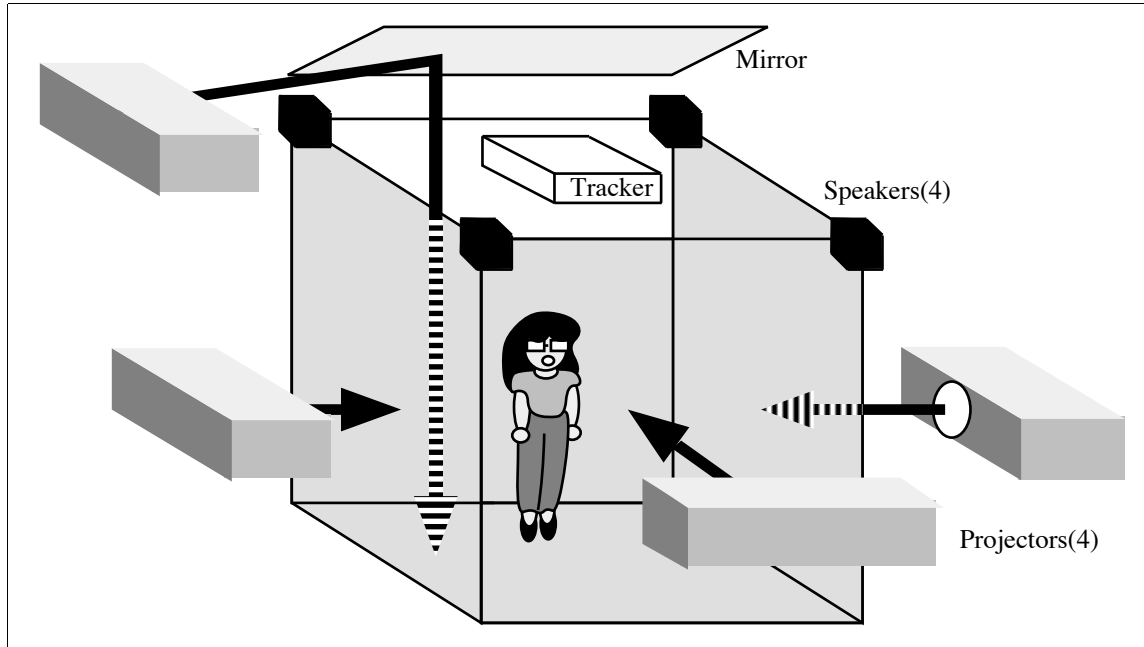


Figure 2: The CAVE

Each instrument placed at a site maintains its own data set in local memory indexed by time. When the user chooses an instrument from the instrument pallet, the preprocessor loads all of its attributes into local memory based on the currently selected time interval. All of these attributes are loaded into local memory before the instrument is placed at a specific site. When the instrument is placed at a specific site the data further partitioned using the selected site. Information on the selected site is kept in local memory, the other information is discarded. With the data values in local memory indexed by time, measurements taken at different time rates can be related.

As the user recreates the experiments the SANDBOX retrieves the appropriate information from the appropriate source. Some instruments are linked to experimental data in the database itself; some are linked to graphical, and other experimental information stored outside the database; some are connected to meta-data stored outside the database.

## 4 Implementation

We implemented a prototype of the SANDBOX in C and GL [9] using the CAVE [2] virtual reality theatre at the Electronic Visualization Lab at the University of Illinois at Chicago. The CAVE is a projection based virtual reality system. The user enters a 10 foot by 10 foot room where images are projected onto the 10 foot high walls. When the user dons a pair of lightweight LCD shutter glasses, the projected images fill the room and surround the user. The user is given the freedom to move around the room reasonably unencumbered. Since they can see their own body, users have a true sense of being in the virtual experiment. To interact with the virtual objects in the CAVE, the user carries a physical three button wand. See Figure 2.

Given the large volumes of data that may need to be transferred between the database and the visualization system, the CAVE has a great advantage over head mounted display technology. During prolonged data access the virtual environment freezes. This can cause the user of a HMD to become disoriented as the movement of their head is no longer reflected by a

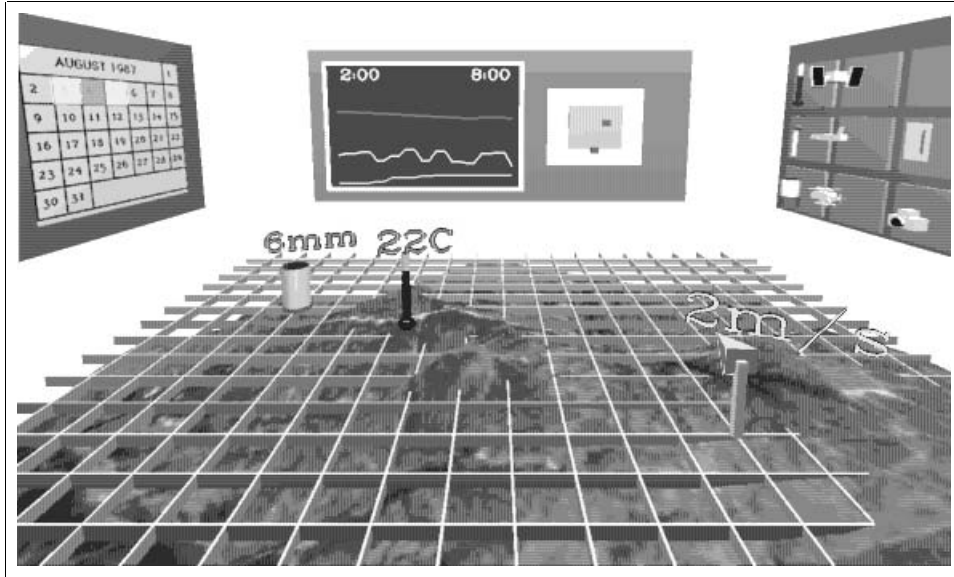


Figure 3: Overview of the Interface

change in their 3D view. In the CAVE, however, while the virtual environment is still frozen, users can see their bodies, and the surrounding CAVE, thereby reducing disorientation and vision induced nausea.

We implemented the SANDBOX on a subset of NASA's FIFE scientific database. The objective of the ISLSCP (International Satellite Land Surface Climatology Project) is to develop techniques to determine surface climatology from satellite observations. FIFE (First ISLSCP Field Experiment) was undertaken at a 20km by 20km square site near Manhattan, Kansas in 1987 and 1989. Its purpose was to gather enough data to allow the creation and testing of models to develop these techniques [19].

120 gigabytes of data was collected (300 megabytes textual data, the rest image data including satellite photos and photographs taken on the ground.) The textual data fills 100 tables in a relational database. Each experiment is given its own table with the attributes containing the numerical data collected during that experiment. These tables typically have 10-30 attributes each, and over 100,000 rows. The textual data is currently available on-line. A subset of the textual and graphical data has been made available on a set of five CD-ROMs from NASA.

In our reenactment of the FIFE field experiments, the user is surrounded by an elevated 3D plane showing the 20km by 20km square site, a pallet of instruments to choose from on the right wall, and a calendar to choose dates from on the left wall. The plane initially shows an enhanced satellite view of the experiment site showing roads, and rivers.

Figure 3 shows an experiment in progress in the SANDBOX. The user has placed a water beaker, a thermometer, and a windsock into the virtual environment. The water beaker shows 6 millimeters of rainfall has collected at its site. The thermometer shows a temperature of 22 degrees Celcius at its site. The windsock shows a northerly wind blowing at 2 meters per second at its site.

Figure 4 shows a close-up view of the instrument pallet. Some of the instruments (thermometer, wind sock, and water beaker in the left hand column) are linked to attributes in the tables of the database. Some (LANDSAT Satellite, airplane, helicopter in the center column) are linked to graphics files stored outside the database [17]. Some (notepad, camera, in the right hand column) are linked to meta-data stored outside the database [18]. The instruments

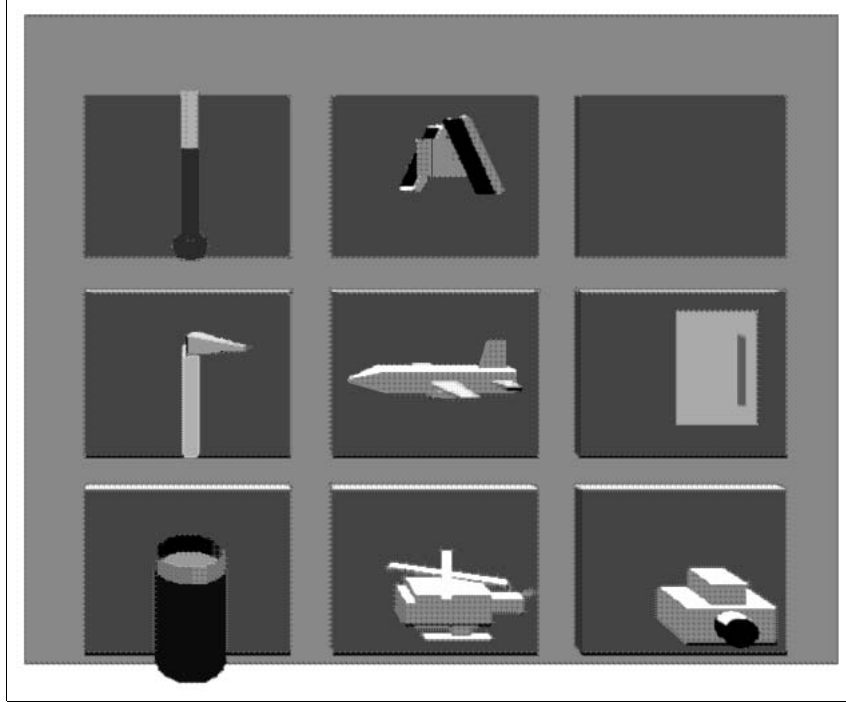


Figure 4: The Tool Pallet

on the pallet are animated (the beaker fills with water, the solar panels on the satellite spin, etc.) to improve their recognizability. The instruments have obvious affordances [12]. All temperatures are measured with the thermometer, all rainfall amounts are measured with the beaker, no matter how they are stored in the database.

The user chooses an instrument by moving the wand to the appropriate instrument on the instrument pallet, pressing a button on the wand, and carrying a three dimensional copy of the instrument off the pallet. All of the sites where the user can place that instrument are then highlighted on the 3D plane. The user places an instrument at a site by moving the wand (and the virtual instrument) over to the selected site and pressing a button on the wand. This gives the user a very natural method of placing instruments as the user can literally pick up an instrument, carry it over to a site, and place it there. The user directly manipulates the objects in the virtual environment [14].

In our current implementation the user can only place a virtual instrument at any or all of the sites where the corresponding real instrument was placed during the actual experiment. In future versions we plan to remove this restriction and allow the user to place the instrument anywhere. The SANDBOX will then interpolate the values at that site from the values at the actual sites stored in the database.

While the amount of data in the database is very large, the number of sites for each experiment is typically small (usually around 10 to 15 sites), giving sparse coverage of the total experiment area. Because of this, conventional visualization techniques can not be employed (i.e. there is not enough information to draw meaningful colour mapped surfaces over the 3D plane.)

Once the instrument is placed at a site, it begins to operate. The mercury level in the thermometer rises and falls with the temperature. The water level in the beaker rises and falls with the rainfall. The orientation of the windsock changes with the direction, and speed of the wind. This allows the user to see how the measurements inter-relate (e.g. the mercury

level dropping in a thermometer as a beaker begins to fill.) If the user needs to see a record of how the values change over time, the values can be displayed in a graph on the front wall. If the user requires quantitative numeric, as well as qualitative graphical values, they can be displayed above each instrument. As with actual instruments, the user can change the settings on the virtual instruments (e.g. the maximum and minimum temperatures displayed by the thermometer, or the units those temperatures are displayed in.)

The user can alter how the 3D plane is displayed. The user can enlarge or shrink the plane. The entire plane can be displayed within the CAVE, or parts of it can lie outside the walls of the CAVE, depending on whether the user wants an overview of the entire experiment, or a close-up view of a certain area. The user can turn the grid lines on to break the plane up into kilometer square blocks, or turn them off to get a better view of the landscape. The user can raise or lower the plane. The plane could be placed on the floor to look down at it from high above, or at waist height to get a better sense of the topography, or above the user's head to look at the terrain from below.

When the user grabs the satellite instrument and places it in the sky above the plane, the user can choose which band to view the 20km by 20km landscape in. The user can choose to see the landscape in visible light, infra-red, or the 5 other LANDSAT bands. LANDSAT photographs from the database are mapped onto the elevated plane. In the actual scientific database the user must refer to a site using its site ID number. In our system the user can see where the sites are located. If the user wishes to measure the temperature near a river, or at high altitude, or where the satellite shows lots of activity in the infra-red, the user can see exactly where to place the instrument by looking at the plane. The graphical information is integrated with the numeric information. In the actual scientific database the user would have to integrate this information manually.

The user views textual meta-data (e.g. site information, notes) with the notepad, and graphical meta-data (e.g. photographs taken at a site) with the camera. If the user wishes to see site information such as longitude, latitude, and elevation, she grabs the notepad instrument and places it at a site. The textual information then appears above it. If the user wishes to see a photograph taken at a site, she grabs the camera instrument and places it at a site. The picture of that site then appears above it. The meta-data is integrated with the numeric information and the graphic information. In the actual scientific database the user would have to integrate this information manually.

Figure 5 shows a close-up view of the calendar. The user selects days from the calendar by clicking on them with the wand. The current virtual day cycles through the set of selected days. A virtual sun orbits the plane rising out of the east each morning at 6am, and setting in the west at 6pm. As the sun dips below the western horizon, the moon rises in the east and continues to orbit the plane opposite the sun. While the Earth's moon does not exhibit this behavior, it gives the user better control over their virtual environment. By grabbing the sun or moon with the wand, the user can adjust the behavior of time, speeding it up, slowing it down, stopping it, or reversing it. The sky changes color as the virtual day progresses, changing from black at midnight, to purple at dawn, to bright blue at noon, to purple again at dusk, and back to black at midnight. This gives the user additional temporal feedback.

## 5 Reaction

Our initial testing in the CAVE suggests that users find this paradigm to be very natural. Picking up and placing instruments appears to be very easy and intuitive. Unfortunately this 'realism' can be physically tiring. Raycasting may be a more appropriate (though slightly less realistic) method of making selections with the wand. With raycasting, a virtual beam emanates from the wand and the user can select anything in the path of the beam. This



| AUGUST 1987 |    |    |    |    |    | 1  |
|-------------|----|----|----|----|----|----|
| 2           | 3  | 4  | 5  | 6  | 7  | 8  |
| 9           | 10 | 11 | 12 | 13 | 14 | 15 |
| 16          | 17 | 18 | 19 | 20 | 21 | 22 |
| 23          | 24 | 25 | 26 | 27 | 28 | 29 |
| 30          | 31 |    |    |    |    |    |

Figure 5: The Calendar

allows the user to remain stationary, making selections at a distance, and has been shown to reduce arm fatigue [6, 15].

Some users have found the calendar and tool pallet to be uncomfortably high. These currently extend from four to six feet off the ground. As discussed previously, we allow the user to change the height of the 3D plane - initially positioned 3 feet above the floor. In future versions we will allow the user to change the height of the entire environment to make it more comfortable for users of different heights.

Originally we positioned the calendar, graph, and instrument pallet just inside their respective physical walls. This gave the users a visual boundary to the environment so they would not accidentally walk into the CAVE walls. Unfortunately, the tracker's accuracy drops off near the walls forcing us to move all three of them one foot into the CAVE. This reduces the user's effective physical work area to 8 by 8 feet.

Some users have expressed an interest in having more sophisticated graphing tools, and the ability to place combinations of instruments (e.g. a beaker and a thermometer) simultaneously. For the former we are looking into interfacing existing visualization tools with the SANDBOX, and for the latter we plan to allow users of future versions to create their own instrument 'packages.'

## 6 Conclusions and Future Work

In this paper we have proposed, and implemented a prototype interface to scientific databases based on experimentation. This interface allows an investigator to deal with familiar instruments rather than unfamiliar query languages, hiding the scientific database from the user. The interface allows the investigator to browse through numeric data, graphical data, and meta-data without concern for where that data is retrieved from.

The SANDBOX encourages experimentation. The user is not sitting at a terminal, typing in queries and receiving columns of numbers as a reward. The user is participating in the

experiment, placing instruments, choosing days, and watching the instruments come to life.

We are currently enhancing our implementation in the following ways: 1) Adding more instruments, and allowing access to larger portions of the database, 2) Allowing multiple investigators using CAVEs at geographically distant sites to cooperate on setting up a virtual experiment, 3) Creating a 3D desktop version of this interface to compare its effectiveness to the current virtual reality version, 4) Investigating various access methods to decrease the data retrieval time from the database, and 5) Adding sound to the environment.

The SANDBOX will be on display in the VROOM of SIGGRAPH '94.

## Acknowledgments

The authors gratefully acknowledge the assistance of Narendra Goel, John Norman, and Don Strebel for their insight into the FIFE database; and the members of the Electronic Visualization Laboratory for their assistance with the CAVE.

## References

- [1] C. Ahlberg, C. Williamson, and B. Schneiderman. Dynamic queries for information exploration: an implementation and evaluation. In *Proceedings of SIGCHI '92*, pages 619–626, May 1992.
- [2] C. Cruz-Neira, D. Sandin, and T. DeFanti. Surround-screen projection-based virtual reality: The design and implementation of the CAVE. In *Computer Graphics (SIGGRAPH '93 Proceedings)*, volume 27, pages 135–142, Aug. 1993.
- [3] R. Earnshaw, M. Gigante, and H. Hones. *Virtual reality systems*. Academic Press, San Diego, 1993.
- [4] J. French, A. Jones, and J. Pfaltz. Summary of the final report of the NSF workshop on scientific database management. *SIGMOD RECORD*, 19(4):32–40, Dec. 1990.
- [5] Y. Ioannidis, M. Livny, and E. Haber. Graphical user interfaces for the management of scientific experiments and data. *SIGMOD RECORD*, 21(1):47–53, Mar. 1992.
- [6] R. Jacoby and S. Ellis. Using Virtual Menus in a Virtual Environment, in Implementation of Virtual Environments. In *Course Notes 9, SIGGRAPH '92*, pages 12.1–12.9, July 1992.
- [7] W. Kim. Object-oriented approach to managing statistical & scientific databases. In *Proceedings of the Fifth International Conference on Statistical & Scientific Database Management*, Charlotte, North Carolina, Apr. 1990.
- [8] M. Kuntz and R. Melchert. Pasta-3's graphical query language: Direct manipulation, cooperative queries, full expressive power. In *Proceedings of 15th International Conference on Very Large Data Bases*, pages 97–105, Amsterdam, Holland, 1989.
- [9] P. McLendon. *Graphics Library Programming Guide*. Silicon Graphics, Inc., Mountain View, California, 1991.
- [10] Z. Michalewicz, editor. *Statistical and Scientific Databases*. Ellis Horwood, 1991.
- [11] J. Nielsen. *Hypertext and Hypermedia*. Academic Press, San Diego, 1990.
- [12] D. Norman. *The Design of Everyday Things*. Doubleday, 1988.

- [13] G. Robertson, S. Card, and J. Mackinlay. Information visualization using 3d interactive animation. *Communications of the ACM*, 36(4):57–71, Apr. 1993.
- [14] B. Schneiderman. Direct manipulation: a step beyond programming languages. *IEEE Computer*, 16(8):57–69, Aug. 1983.
- [15] C. Shaw, J. Liang, M. Green, and Y. Sun. The decoupled simulation model for virtual reality systems. In *Proceedings of SIGCHI '92*, pages 321–328, May 1992.
- [16] G. Stephenson. Knowledge browsing - front ends to statistical databases. In *Proceedings of IVth International Working Conference on Statistical and Scientific Database Management*, Rome, Italy, June 1988.
- [17] D. Strebel, D. Landis, J. Newcomer, D. van Elburg-Obler, B. Meeson, and P. Agbu. Collected data of the first ISLSCP field experiment, volume 2: Satellite imagery, 1987-1989. Published on CD-ROM by NASA, 1992.
- [18] D. Strebel, J. Newcomer, D. Landis, J. Nickelson, S. Goetz, B. Meeson, P. Agbu, and J. McManus. Images derived from the first ISLSCP field experiment, volume 5: Images derived from satellite, aircraft, and geographic data. Published on CD-ROM by NASA, 1993.
- [19] D. Strebel, J. Newcomer, and J. Ormsby. Data management in the FIFE information system. In *Proceedings of International Geoscience and Remote Sensing Symposium*, pages 42–45, Vancouver, Canada, July 1989.
- [20] G. Wyszecki and W. Stiles. *Color Science*. John Wiley and Sons, New York, 1982.