SAND REPORT

SAND2006-061506150294
Unlimited Release
Printed December 2005

LDRD Final Report: Improving Human/System Interactions in Systems-of-Systems

Stephenson Tucker, Carl E Lippitt, Mark Gardener, Kenneth L Summers, Gregg D Whitford, Derek H Hart, Michael J McDonald, Fred J Oppel III, and Mary Fendley

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185 and Livermore, California 94550

Sandia is a multi-program laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

Sandia National Laboratories

"...exceptional service in the national interest."

Issued by Sandia National Laboratories, operated for the United States Department of Energy, by Sandia Corporation.

SAND2005 – 0294
Unlimited Release
Printed October 2005

LDRD Final Report:
Improving Human/System Interactions in Systems-of-Systems

Stephenson Tucker, Carl Edward Lippitt and Mark Gardener
Cognitive and Exploratory Systems

Michael J. McDonald, Fred J. Oppel III,
Kenneth L. Summers, Gregg D. Whitford and Derek H. Hart
Intelligent Systems Principles

Mary Fendley
Cryptography and Information Systems Surety

Sandia National Laboratories
PO Box 5800
Albuquerque, NM 87185-1004

Abstract

The main premise of this LDRD is to allow a decision-maker that has little familiarity with modeling and simulation to be able to interact with simulations in a naturalistic manner similar to ways in which humans communicate about ideas to others. We report on a user-centric Human-System Interface (HIS) prototype, based on the taxonomy of observed human interactions integrated with 3-D tangible interface coupled with voice and gesture recognition. We have demonstrated a "Tangible Table," based on gesture and speech, as a means of allowing humans to interact with simulations of System-of-Systems (SoS). A number of applications are limited primarily by their HSI. Unlike machines, humans are adept at conveying information between themselves using a variety of techniques, including speech, gesture, using props (tangibles), and drawing. By coupling technologies that mimic these interactions, we are developing interface technologies for complex systems that are both intuitive and user-centric in design. Our focus was on the development of an HSI for Modeling and Simulation (M&S), especially analytic tools. We used Sandia's Umbra simulation framework and added Object Oriented containers to support the easy creation of hierarchical systems representations; we implemented the concepts of Architectural Description Languages utilizing XML, based on Carnegie Mellon University's ACME.

Keywords

Tangible Interface, Human Computer Interface, Usability

Acknowledgment

Carl Edward Lippitt, 48, the technical lead on this LDRD, died Wednesday, August 31, 2005. He is survived by his wife, Diane, daughters, Casey and Caitlyn and son, Christopher. He was a graduate of the University of Missouri in Rolla. He worked for Sandia National Laboratories from 2001 to present. Carl passed away after the successful completion of this LDRD's final demonstration, but before the start of this report. Carl embodied much of the technical expertise on this LDRD; He will be missed mostly for his wry sense of humor, sincere kindness and inquiring mind.

Carl kept a daily blog of his thoughts as he studied the Bible. The following is the final portion of the final entry Carl logged on the day of his heart attack:

> *"If the only time we have is today, the now, is it not wisdom to consider both the future and the past as the Lord's? Rather than covet what we used to have, give thanks for what*

remains and look with hope for that which is promised to all who are eagerly waiting the return of the king."

Introduction

The main premise of this research is to allow a decision-maker that has little familiarity with modeling and simulation to be able to interact with simulation in a naturalistic manner similar to ways in which humans communicate about ideas to others. When properly approached, the simulation process can provide insights into a system that no external view can. The push to utilize Modeling and Simulation (M&S) for complex analytics has created tools that are useable only by specially trained modelers. To date, the capability of most M&S applications is underutilized due to lack of tools that enable effective management by humans. This paradigm has the end-user of the analyses—the person posing the questions—needing assistance from modelers to have their questions answered. The research presented here was inspired by the observation that often there is a detrimental de-coupling of the decision maker who wants a question answered by a simulation and the answers that can be obtained in the process of generating an appropriate simulation. Our research concerns how to involve the end user in the simulation process to aid in their understanding of a system.

To meet this goal, we developed a user-centric HSI prototype, based on the taxonomy of observed human interactions integrated with a 3-D tangible interface coupled with speech and gesture recognition. For the M&S framework, we used Sandia's Umbra, to which we needed to add a better means of representing hierarchical systems, a tangible interface and speech recognition. Our desire was to develop a digital version of the military sandbox, where a collaborative group could pickup physical objects, talk about their representation or function in a "human-to-human" like fashion, have the simulation understand the interaction and show the results. The tasks fell into two categories:

- Adding Object-Oriented Containers to Umbra to support the easy creation of hierarchical systems representations.
- Implementing the tangible/speech interface to Umbra

After an investigation of Architectural Description Languages as a means of representing hierarchical systems, we chose to implement the concepts of Carnegie Mellon University's ACME utilizing the eXtensible Metadata Language (XML) (McDonald and Rigdon, 2005). This report will cover our work on the tangible/speech interface to Umbra.

Definition

A widespread definition of tangible interfaces is: *"Generally graspable and tangible interfaces are systems relating to the use of physical artifacts as representations and controls for digital information. A central characteristic of tangible interfaces is the seamless integration of representation and control, with physical objects being both representations of information and as physical controls for directly manipulating their underlying associations. Input and output devices fall together."* Ullmer and Ishii (2000)

Current work

Hiroshi Ishii, of the MIT Media Lab, has done extensive work with a specific type of interface that he calls a tangible user interface. To understand the difference between the traditional graphical user interface and the tangible user interface we can look at their differing properties. Traditional graphical user interfaces (GUIs) cause us to shift our attention between interaction with the device, and the information display. This can lead to confusion and delay (Mazalek et al., 2002). GUIs make a distinction between input devices/controls (keyboard, mouse) and output devices/representations (monitors, head mounted displays). Tangible user interfaces (TUIs) integrate representation and control (Ullmer and Ishii, 2000). The following are properties of TUIs laid out by Ullmer and Ishii (2000):

- Physical representations are computationally coupled to underlying digital information (model).
- Physical representations embody mechanisms for interactive control (control).
- Physical representations are perceptually coupled to actively mediated digital representations.
- Physical state of tangibles embodies key aspects of the digital state of a system.

Ishii's published work addresses a future need for usability evaluation for human-computer interaction, but does not specifically address these issues. Some of the future concerns he mentions are: physical scale and situatedness, cognitive engagement and distance, general versus special-purpose approaches, and what makes for a good tangible interface. Shneiderman's three principles of direct manipulation are for GUI's, but the first one—continuous representation of the object of interest is especially applicable to Tangible User Interface tangibles. This addresses several of the usability concerns listed for this project.

Ullmer and Ishii (1997) define the Tangible User Interface (TUI) as a user interface employing using physical objects, instruments, surfaces, and spaces as physical interfaces to digital information. More specifically, they have been working on a project called the metaDESK, which is a graphically intensive system driven by interaction with graspable physical objects. An important human-computer interaction design principle that the metaDESK addresses is seeing and pointing vs. remembering and typing. Their TUI uses tangible bits, with the goal to bridge the gap between cyberspace and the physical environment by making digital information tangible to the user (Ullmer and Ishii, 1997). They are able to do this through interactive surfaces, coupling of bits and atoms, and using ambient media (sound, light, airflow). This approach, much like the tangible pucks used with this project's interface, has a strong focus on graspable physical objects as input.

Planar Manipulation Device (PMD)

During SIGGRAPH '03, Dr. Dan Rosenfeld of New York University's Center for Advanced Technologies NYU/CAT, was displaying the Planar Manipulation Device (PMD) (http://cat.nyu.edu/PMD). The PMD is a computer peripheral which allows multiple objects to be moved and sensed by application programs running on a host PC. Small motorized platforms (vehicles) are used to move physical objects used in the interaction or simulation. The system can sense when objects are moved by participants and allows multiple participants to interact simultaneously with the application.

We later visited NYU/CAT for further discussions and demonstrations. First, a discussion of why we did not use the PMD; the tracking hardware utilized was based on an optical Position Sensing Device (PSD) which generates an analog signal based on the position of the image of a light source (in this case an LED mounted on the tangible) when imaged onto the PSD. The analog signal was converted to a digital signal using an 8-bit Analog-to-Digital (A/D) converter. Over an approximately 3 foot by 3 foot surface, this yielded a minimum resolution of position (X, Y (Single Bit)) of greater than 0.1 inch. Studies of computer mice show the need to have resolutions of 0.01 – 0.001 inch, for usable pointing. While this is a solvable problem, the lack of maturity was worrisome for a larger integration effort, such as this LDRD.

While we chose not to utilize the PMD, it has many unique features that merit continued attention. Specifically, the ability of either the human or the simulation to move the tangible is intriguing. There is still a large, pending usability question about what happens when entities move in virtual space, without a corresponding movement of the tangible. Our implementation could support mobile tangibles. This is an area that warrants follow-on work.

ARToolkit

Our initial proposed implementation was with a software library called ARToolkit (ARTK). ARTK is primarily developed by Dr. Hirokazu Kato, Osaka University and supported by the Human Interfaces Technology Laboratory, Washington University (http://www.hitl.washington.edu/artoolkit/). We had gained significant experience with ARTK in a previous project, and initially it seemed to be a good choice for our tangible interface.

ARTK is a software library primarily for building Augmented Reality (AR) applications. These applications involve the overlay of virtual imagery on the real world. ARTK uses computer vision algorithms to track symbolic markers in a video scene. Its video tracking libraries calculate the real camera position and orientation in Six Degrees of Freedom (6-DOF) relative to physical markers in real time. Knowing the camera frustum in 6-DOF allows 3D virtual objects to be placed into the scene relative to the correct viewpoint.

Initial thinking was that a fixed camera could track sets of these markers for the direct manipulation of the simulation. Additionally, users would see rich 3D representations of information on the appropriate marker. Since ARTK runs on a standard laptop with an inexpensive Universal Serial Bus (USB) camera, the low cost of hardware is appealing.

The ability to display 3D virtual objects enables the capability to exhibit more intuitive and rich data representations; however, the user must wear a Head Mounted Display (HMD) to immerse them into the correct spatial orientation. Experience has shown that novice users are reluctant to use HMD and are much more comfortable with flat 2D displays.

While ARTK runs very well in simple demonstrations, it did not scale well for tracking multiple objects. Latency increased with multiple markers to the point of extreme distraction. Even in simple demonstrations, ARTK consumes all of the computer resources, which causes latency issues in Umbra. ARTK did not prove to be a viable solution for this LDRD, its demonstration led to a separate project – the "Augmented Reality Training System" or ARTS, a Close Quarters Combat training simulator, funded by DOE/NNSA and another Work-for-Others (WFO) customer.

Final Implementation

With the experiences we gained, we developed a set of requirements for the final implementation of our TUI. These included:

- Support for a multi-modal (Gesture/Speech) interface paradigm
- Support for a wide variety of gestures
- The ability to display information via separate mechanisms onto the flat surface and the tangibles. This could be implemented by a monitor, i.e. LCD or separate external projection, bottom projection for the surface and top projection for the tangibles.
- The ability to support 3D Augmented Reality representations
- Low latency (less than 20 mS)
- High Spatial Resolution (0.01-0.001 Inch)
- Wireless Tangible Devices (RF or IR)
- The ability to handle multiple tangibles without an increase in latency (at least 10 tangibles)
- Support for the DOD High Level Architecture (HLA). HLA is a general-purpose architecture for distributed computer simulation systems. Using HLA, computer simulations can communicate to other computer simulations regardless of the computing platform.

During our research we did not become aware of any work that supported even most of this list. However, we found a separate technology that lent itself well to our application. This technology came from a company named PhaseSpace (http://www.phasespace.com).

PhaseSpace Optical Motion System

PhaseSpace is a manufacturer of optical motion capture equipment. Motion Capture, or MOCAP, is a technique of digitally recording the movements of real things – usually humans. It originally developed as an analysis tool for biomechanics research, but has grown increasingly important as a source of motion data for computer animation. In this application, it has been widely used for both cinema and video games. In the movie series "The Lord of the Rings," the character "Gollum" was created by computer animation from MOCAP. At the time of our studies, the PhaseSpace motion digitizer had the best performance attributes (such as system resolution, speed and range) of existing systems.

Figure 1 – Phasespace System

Operationally, the PhaseSpace system (figure 1) uses multiple "cameras" to locate multiple LEDs. In their parlance, PhaseSpace's "camera" refers to an optical sensor with (2) orthogonal linear CCDS and internal processing (FPGA and DSP). Each camera outputs the polar angle (ϕ) and azimuth (θ), of each LED, in spherical space relative to the X, Y plane of the camera. A computer gathers this data from multiple cameras and triangulates the 3D position (in physical space) of each LED from these sets of vectors. Each LED is usually visible from multiple cameras. For MOCAP, the system (12 – 24 cameras) is arranged in a ring (figure 2), and an actor (figure 3) moves about within the ring.

Figure 3 – Actor with LEDs
(Note captured data on screen)
Courtesy of Phasespace

Figure 2 – Typical MOCAP Ring

The PhaseSpace system is modular in design and met or exceeded all of our applicable performance criteria:

- Up to 120 individually distinguishable LED targets
- 480Hz frame rate for all targets
- Basic spatial resolution (X, Y) - 3600 x 3600

Conceptual Design

We developed a concept for implementation shown in figure 4 based on the PhaseSpace system. We would develop tangible “pucks” that could be placed on top of a commercially available rear projection screen (translucent) that was built into a table frame. The “pucks” (similar in size to a hockey puck) would contain (3) downward-pointing LEDs, RF wireless electronics and several “gesture” mechanisms. The frame would also support a large front-surface turning mirror, a High Definition (HDTV) projector and 4 PhaseSpace cameras. The projection system would work much as most standard systems. Since the projection surface is translucent, when the pucks are placed on the surface they are visible to the cameras. Not shown in figure 4 is the concept for the top projection system. A separate projector (ceiling-mounted pointing down) projects information onto the top of the puc'

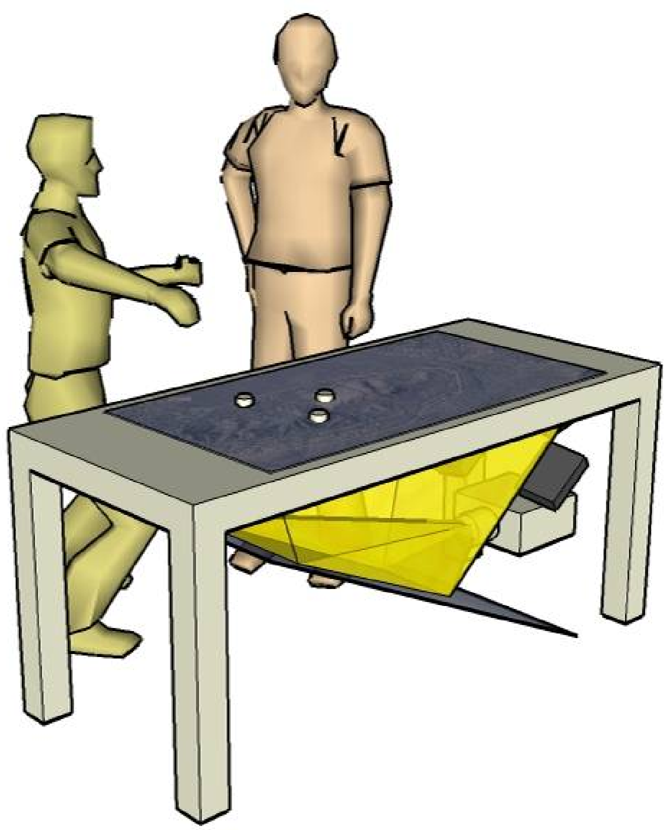


Figure 4 – Conceptual Table Layout

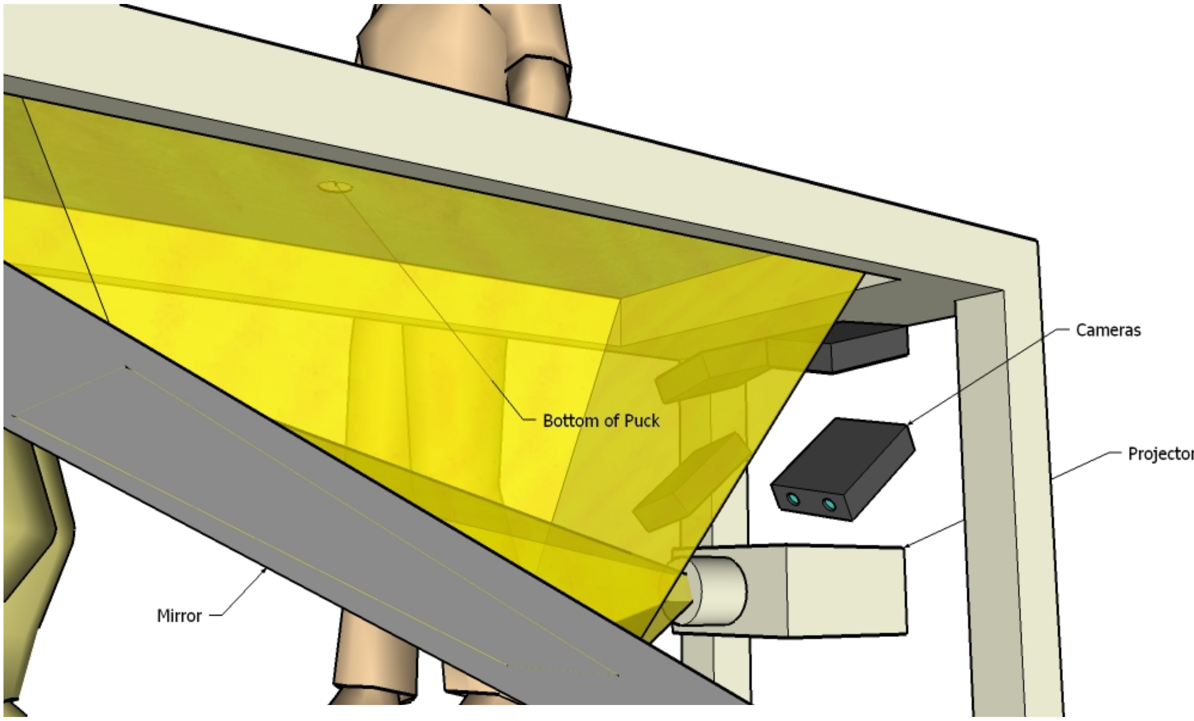


Figure 5 – Conceptual Internal Layout

F

igure 5 shows more on the conceptual internal layout of the table. The table dimensions are based on ergonomic reasoning: the height (36") is waist high for a median standing person, which allows someone to easily bend over the table to reach a puck on the other side. The width (32") is based on the distance a median standing person can easily reach across. The boarders along these sides are 1" which allows a maximum screen size of 30". Since the projector is HDTV with an aspect ratio of 16:9, the screen length is 53". The length of the table (80") allows pucks to be set aside on the table ends. The turning mirror (front surface silver) is required to accommodate the throw distance of the projector and cameras.

Two cameras must see each LED to triangulate its position. At the current throw distance, each camera images a 30" x 30" area of the table surface. Since the projection surface is 53" x 30", one set of (2) cameras images one end of the table, and another set of (2) cameras images the other. There are several inches of overlap that is seen by all four cameras. This arrangement creates a single-bit resolution of less than 0.01".

In designing the pucks, our concept was to use (3) downward-pointing LEDs arranged in an equilateral triangle. This idea allows the pucks to be tracked in both position and rotation. While not used in this LDRD, this design also accommodates full 6-DOF tracking of the pucks, which is useful in other forms of gesture (pointing, 3D drawing…). Ergonomically, we found a puck diameter of 3" and height of 1-2" to be easily and intuitively manipulated.

Tangible Puck Development

At the beginning of development, PhaseSpace had just released a RF wireless driver for chains of LEDs to be worn by actors in MOCAP. The device is belt worn and ~ 4"W x 8"L x 2"D. The actor in figure 3 is wearing one behind his back. Additionally, at the time the system only supported one wireless device at a time.

We worked with PhaseSpace to generate a new set of electronics and form factor to fit our puck design. The result is shown in figures 6-8. The results include:

- 3" Diameter and 1" Tall
- Support for up to 40 independently tracked pucks
  - 480 Hz
  - No increase in latency with number of pucks
- Spatial resolution ~ 0.01" (single bit)
- Rechargeable batteries
- Gestures include:
  - Position
  - Rotation
  - Contact with projection surface
  - Pressure switches mounted on top (push) and around periphery (squeeze)

Additionally, we worked with PhaseSpace to incorporate the electronics into a gesture glove that included LEDs and shape sensing tape. The system would be tracked by a separate set of cameras mounted above the tangible table. The initial idea as that the user would make a gesture such as "kids playing guns" and point at the puck of interest. The work on the gesture glove was terminated when interfacing two separate complete PhaseSpace systems proved problematic. This will be resolved in future versions, but not before the end of this LDRD. The gesture of touching or picking up a puck to indicate attention has proven to work very effectively. The speech recognition system currently allows the user to interact with the last puck that has been handled.

Figure 6 – Puck Top
(Note Pressure sensing devices on top and around sides

Figure 7 – Puck Bottom
(Note LEDs (3) on PCB and contact switch)

Figure 8 – Puck internals

Speech Recognition System

The tangible pucks provide an intuitive way for humans to use gesture to interact with a computer. However in human - human interactions we also use speech to describe concepts, problems, behaviors and entities. Our design includes a Speech Recognition (SR) component. Since computer understanding of free speech is difficult, we decided to simplify our SR interface by using voice menuing technologies, with a video based "teleprompter." Limiting the number of word/phrases choices also substantially improves accuracy and decreases latency. Voice menus are easily described in XML, our data description language.

We decided on a commercially available SR product from Lumenvox. (http://www.lumenvox.com). In use, the user says one of several phrases displayed, and can thereby drill down through the available options in hierarchical menus.

Our complete system was located in a very noisy laboratory. Background noise is a problem for our SR system – usually one would have to repeat oneself or suffer long latency. We tried several different microphones; directional “room” microphones were very problematic, head mounted “boom” microphones were less so. Our primary consideration for using Lumenvox was their claim that they were developing a Linux version of their “Speech Platform.” The other portions of our system were implemented in Linux. Other commercial products may be better.

System Design

The TUI system that we designed and developed is diagrammed in figure 9.

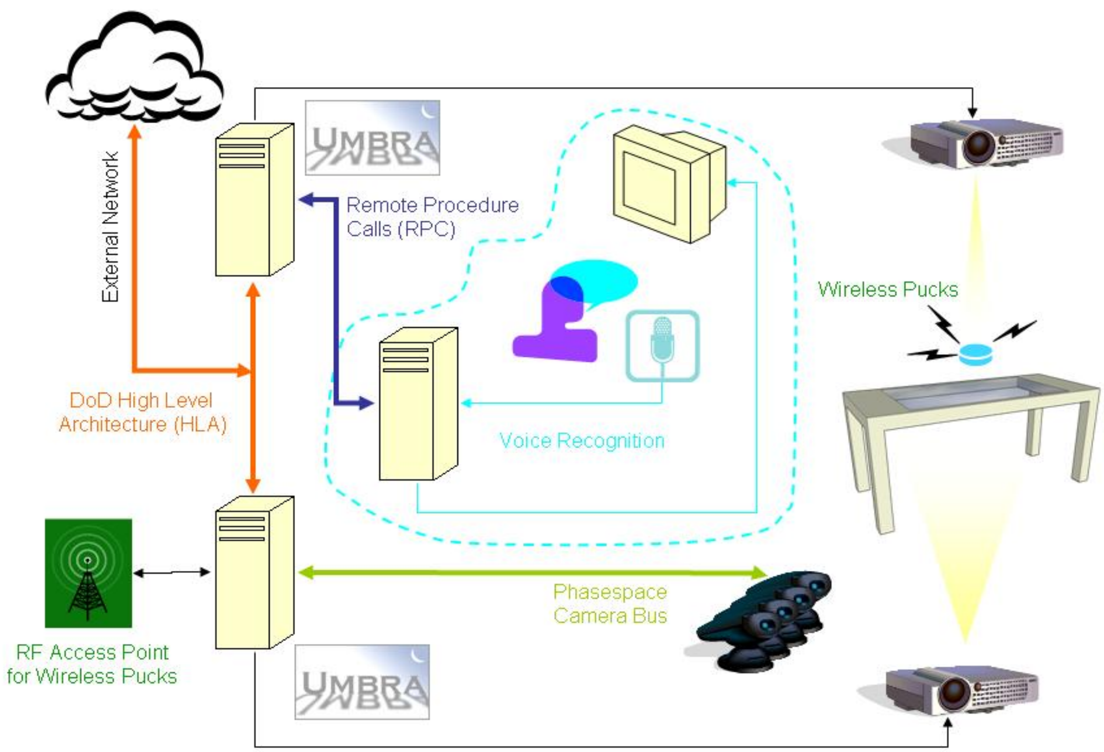


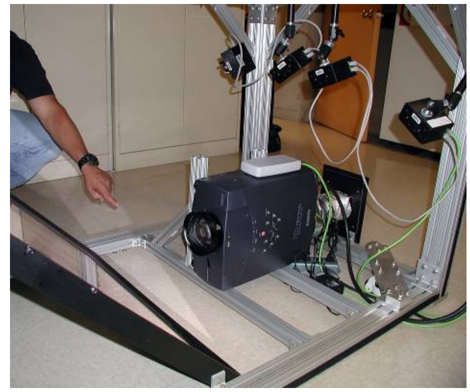
Figure 9 – System Diagram

The TUI system is comprised of three subsystems. These are described separately:

- The bottom system is responsible for:
    - Umbra simulation space
        - Understands virtual world coordinates
        - Projects simulation representations onto bottom of the table
    - PhaseSpace hardware support
        - Supports tangible puck tracking
        - Supports bi-directional RF communications with pucks
        - Synchronization
    - Recognizes gestures with the puck
    - Calibration
- The top system is responsible for:
    - Umbra table space
        - Understands user interactions with pucks
        - Understands user speech commands
    - Interprets user intent in simulation space from table space interactions
    - Coordinate transformations between simulation and table space
- The SR systems is responsible for:
    - Displaying video voice menus
    - Converting speech into commands

These three systems are connected via a network. The top and bottom systems communicate between themselves and other federates via the DOD High Level Architecture (HLA). The SR system uses standard Remote Procedure Calls (RPC) to communicate via the network with the top system.

Figure 10 shows the completed table. Figure 11 is from behind the bottom projector, looking at the mirror. Note the LEDs from (2) pucks are visible on the table above. Figure 12 shows the bottom projector and (4) cameras. Figure 13 shows the TUI in use during our final demonstration. Note the puck IDs projected from above, menus associated with each puck projected from below, and the SR menu screen (vertical LCD).

Figures 10, 11 & 12 – Completed Table and
Internal views

Final Demonstration

We developed a scenario for our final demonstration (figure 13) that emphasized the types of complex SoS problems that are characteristic of today's national security issues. We wanted to underscore the benefit of allowing the human to interact with simulations to gain insights and branch into other directions based on those insights. Additionally, we wanted to show the benefit of blending human intuition and perception with the computational benefits of simulation.

Figure 13 – Operation during demonstration

The goal for the user in the demonstration was to find an adversary vehicle in a MOUT (Military Operations in the Urban Terrain) terrain using sensors, RF communications and an Unmanned Aerial Vehicle (UAV). We specifically developed the scenario to utilize the more robust model sets that Umbra has, which included vehicles, sensors, communications and UAVs. For a terrain we used a notional MOUT terrain shown in figure 14. Note: for illustrative purposes the figure shows the terrain in an isometric view. In most cases (demonstration included), the view on the tangible is top-down, or Plan View.

Within the terrain, we used Umbra's route planning and following algorithms to have a set of vehicles driving around including (1) friendly (Blue), several neutral (Green) and (1) adversary (Red). The Red vehicle, in a repeating cycle and following several independent routes, entered from outside of town and drove to a building, where it parked inside for various amounts of time and then left town. The Blue and Green vehicles would randomly drive around the MOUT roads. For simplicity, vehicles will be referred to by color.

Figure 14 – Simulation Demonstration Terrain
Note: Sensor viewing frustums

The task of the user was to find/observe Red with the following capabilities/caveats:

- Blue serves as the Command and Control (C2) node for this demonstration. Red can either be directly observed by a sensor mounted on Blue or if Blue has RF communications with another sensor that is observing Red.
- Communications is modeled based on Line-Of-Sight (LOS) between transceivers (nodes). If nodes have LOS, RF range fades at $1/r^2$ - if not LOS, then $1/r^4$. There are nodes on all sensors, the UAV and Blue. RF range is a variable, usually set at the beginning of the demonstration.
- The sensors that are modeled emulate static ground-based video sensors that can see vehicles within their viewing frustums. Variables that the user can set, via the TUI, are: position, orientation (rotation), transmitter power, receiver sensitivity and Field of View (FOV). FOV is set by using the puck dial.
- The UAV flies between (4) waypoints that the user can set. These waypoints are modeled as entities on the terrain; the user sets, via the pucks, the position of the waypoint as well as the elevation (dial) above the waypoint entity. The UAV has two additional functions: it carries a downward facing video sensor and it acts as a communications repeater. Red is detected if:
    - UAV sensor sees Red and UAV has communications with Blue
    - A ground-based video sensor sees Red and has communications with the UAV and the UAV has communications with Blue.

Additionally, the pucks could be used to zoom and pan in the simulation scene.

Typical tradeoff questions the user might explore are:

- How to place the ground-based sensors?
    - For optimal sensing
    - For optimal communications
- What path should the UAV follow?
- Is it better to use the UAV as a sensor or as a repeater platform?
- Are different altitudes better for different purposes?

Usability

The goal of the project was to design a system where users can interact with the system and each other in a natural way. Usability issues are an important aspect of system design to ensure natural user interaction and acceptance. The following are a few of the usability issues that have presented themselves.

- Assignment of object to puck
- Selection point on puck.
- Size of entities
- Control of continuous range (zoom)
- Number of users per puck
- Continuously connect or disconnect user
- Overhead projection onto puck
- Percent of speech versus tangible puck manipulation
- Hook/orientation
- Mapping physical to the virtual world.

Traditional usability evaluation methods

Paternò (2005) discusses how model-based approaches and tools have been used to address important issues for obtaining usable interactive software. He states that task-based approaches are suitable to design user-oriented interactive applications because they effectively and efficiently support users' activities. More specifically, task modeling is a method that builds a model which precisely describes the relationships among the identified tasks. Bowman (1999) describes the advantages of using a testbed for usability evaluations. A testbed is defined as a representative set of tasks and environments. Testbed evaluations combine multiple tasks, multiple independent variables, and multiple response measures to obtain a more complete picture of the performance characteristics of a system.

Challenges facing evaluation of multimodal interfaces

New challenges arise for model-based approaches in natural development, ambient intelligence, and multi-modal interfaces. In multimodal interfaces task performance is influenced by the modality available. Here we have both visual and speech. Vocal channels are suitable for simple or short messages for signaling events, immediate actions, to avoid visual overload, and when the user is on the move within the system. Visual channels are more appropriate for complex or long messages, identifying spatial relations, when multiple actions need to be performed, when interacting in noisy environments, and for stationary users (Paternò 2005).

Ways to provide feedback or input information through these multiple channels need to take into account features of the available platform. An analysis of space of possible design choices can be made through CARE properties (Paternò 2005):

- Complementarity—synergistic use.
- Assignment –specific modality is selected.
- Redundancy—multiple modalities for the same purpose.
- Equivalence—there is a choice of one modality from a set of available ones.

Another method to address the challenges of evaluating multimodal interfaces is to consider them from the point of view of a virtual environment. According to Stanney, et al. (2003) some of the limitations of traditional usability methods for assessing virtual environments include the following. First, point and click interactions are not representative of multi-dimensional object selection that may occur in a multimodal environment. In this project both voice and puck manipulations are used for object selection. Second, the quality of multimodal system output is not comprehensively addressed. These modalities include visual, auditory, and haptic. Third, single user task assessments do not consider virtual environment system characteristics in which two or more users interact in the same environment. Here we are anticipating that more than one user will be interacting with the system at any given time.

Interface considerations that need to be taken into account may include some of the following. A usable virtual environment system needs a flexible underlying model in order to allow natural user interaction with the system. Interaction should be natural, efficient, and appropriate for target users, domains, and task goals (Stanney et al. 2003). There are three categories of interaction: travel, selection, and manipulation. Another aspect to consider is wayfinding, where the user manipulates their point of view to move from place to place. In this case the user can maneuver between points of view from different pucks representing the various entities within the system.

Assessment methods for usability applicable to this project

Stanney et al. (2003) use Multi-Criteria Assessment for Usability for Virtual Environments (MAUVE) for evaluating the usability requirements of virtual environments. MAUVE is an approach where the evaluation goes beyond the mere subjective opinion of the system evaluators. A set of evaluative criteria is developed to guide evaluators through a heuristic assessment of virtual environment usability. This occurs in two stages. In the first, traditional usability heuristics are applied. Second, a multiple criteria decision-making technique prioritizes usability criteria according to the needs of a given application.

This project does not involve a virtual environment per se, but has many of the characteristics of a virtual environment. The following are aspects of MAUVE are that seem to map to the usability concerns of this project.

Interaction usability concerns with navigation within the system need to be able to answer the following questions:

- Is it easy for users to move and reposition themselves in the environment?
- Is a navigational grid or map included for large environments? Do implemented maps adhere to map design principles?
- Is the level of user movement control appropriate for the specific task?

At this point, the user can navigate within the system using the puck; More specifically, through physical movement to change position and rotating the puck to zoom the simulation scene.

Interaction usability concerns with object selection and manipulation should be able to conform to the following:

- Input devices should be easy to use and control (not too sensitive or sluggish)
- Query formation (command or speech input) can be used to assist with object selection methods.
- Selection based on spatial attributes (location, shape, orientation) should be supported via direct manipulation.
- Selection based on temporal, relational or descriptive attributes should be supported via non-direct manipulation.

This goes back to the issues of how much manipulation should be controlled through speech versus tangible puck manipulation, and is directly related to multimodal system output usability concerns. These include visual and auditory input/output. Here a visual display should be seamlessly integrated into users' task activities. The visual and auditory display lag should also be non-cumbersome. Distortions and lag in visual output affect user perception and therefore situational awareness and trust in the system.

These usability issues are addressed by Dybkjaer et al. (2004) through the use of Spoken Language Dialogue Systems. Concept accuracy/error rate is a popular measure for the extent to which the natural understanding captures the key concepts in user input to the system. It is critical for the system to have the ability to understand spoken input and appropriately respond measured in terms of the information returned to the user.

Dybkjaer et al. (2004) propose six Spoken Language Dialogue Systems aspects to be analyzed: speech recognition, speech generation, natural language understanding and generation, dialogue management, human factors, and systems integration. Integrated with these are several key usability issues relevant to this particular system. These include the following. Input recognition accuracy needs to be high (for user confidence), in terms of naturalness of user speech and feedback adequacy. The user needs to be confident that the system has understood the input in the intended way. The system needs to be sufficiently transparent; accordingly, the user needs to be told what actions the system has taken and what the system is currently doing. Naturalness of dialogue structure needs to be considered for ease of use and user acceptance of the system, as users have an expectation to the information/service they should get from the system. The system needs to be able to handle errors adequately and be able to sufficiently adapt to user differences.

Observations

Overall the system performed much as expected, and the feedback from those who tried it was uniformly positive. We also learned several important lessons and found areas for further research.

According to the literature review and observation of the system, there seems to be two main usability issues regarding user acceptance of the system. These are the recognition speed and errors of the spoken system, and the correspondence between the physical and virtual world.

From the first demonstration of the system/table it is very apparent that the system is slow to recognize a speech input command. Whether it is from not recognizing the speech or when a command starts and stops, the result is still the same—the system locks up or is extremely slow to respond. Oviatt (2000) presents a solution of employing a system fusing more than one type of information source to reduce recognition uncertainty.

An important factor in speech input in this system is the issue of multiple users. The table is intended to be used by multiple individuals simultaneously. Putting aside for the time being any issues of multiple person voice recognition, a problem still remains as to where the voice command is to be applied. We used the rule that the last puck "touched" (meaning the puck which last had a pressure sensor activated) was the one to which the voice command applied. In a multiple user environment, this gets confusing as, at the time the voice command is being acted upon, there may be multiple pucks with active pressure sensors. Someone may grab a puck other than the intended one during the voice command, or during the lag between the command and the parsing of that command. These interactions seem quite natural to the participants, and not at all distracting, because body language, memory, and implied possession are all used to differentiate the target, but none of these are available to the computer. More development will be required to more naturally integrate voice commands and puck possession interactions.

Another issue is that of the puck occluding portions of the scene. Each puck cuts a 3" hole from the user's view of the virtual world. This presents problems with object selection and manipulation – the obvious method of selection, where the puck is placed over an object then either a voice command or pressure is used to select it, is infeasible due to this occlusion. Our solution was to create an offset reticle that represents the puck. A crosshair in the reticle makes precise selection and manipulation possible. Since the majority of users are right handed, and will be standing at the "bottom" of the screen, the reticle was placed to the left of the puck, to minimize obscuration by the hand and arm. An alternate solution that was discussed was creating pucks with holes or clear portions in the center.

A mode for using the top pressure sensor that was quickly adopted was to use it like a mouse button. People who have used a mouse tend to gravitate to that model when using a puck, so it was adopted for object selection and menu navigation.

Creation of a new model for interaction requires the creation of new tools. Two tools that were adopted and modified from other areas were the rheostat and the menu navigator. The rheostat is a semicircular rotational tool centered on the puck and projected from the bottom. Radial lines extend past the borders of the puck, with a triangular indicator, much like the large rheostats found in older electronic equipment. The current numerical value of the variable being controlled is displayed on the top of the puck. This tool is used to set continuous, bounded values required by the simulation, such as sensor field-of-view and transmitter power.

Menu navigation is similar to that used in standard hierarchical computer menus. The difference is twofold: the menus are called up by voice command and displayed next to the selected puck, rather than at a fixed location on the screen and, instead of moving a marker like a cursor to the desired selection, the puck is rotated until the desired selection is highlighted. The puck is then "clicked" just like a mouse to finalize the selection.

From a usability perspective, a system that employs a multimodal interface takes advantage of the human's intelligence in decide how to effectively use input modes. The human operator responds in three ways in order to reduce recognition errors. First, they use the mode most likely to be accurate for certain lexical content. Second, the user tends to use briefer, simpler language, thereby reducing the complexity of natural language processing. Third, the user alternates between input modes to aid in error reduction. The modes may be asymmetrical in their reliability, so system development should have alternate modes that complement one another to support the reduction of confusion.

Conclusion

We have described our implementation of a Tangible User Interface, the "Tangible Table." We have successfully demonstrated the Tangible Table as a means of allowing humans to interact with simulations of complex SoS. Our results with the tangible interface (pucks) have met our expectations; however, the speech recognition portion of our system has sufficient latency to have a serious negative impact on "human-like" computer interactions. The tangibles operated well and both confirmed our expectations as well as pointed out areas that need to be examined in the future. Overall, feedback on the system has been very positive.

As the final demonstration was completed at the end of this LDRD, there is still much work to be done in usability. Our implementation has sufficient flexibility to support incorporation of "lessons learned" in future usability studies.

During this LDRD we had multiple interactions with the Integrated Media Systems Center at the University of Southern California (USC/IMSC). Specifically we were interested in their advanced speech and gesture work (http://iris.usc.edu/~icohen) and (http://sail.usc.edu/shri.html). We believe that future work should include attention to improvement of both speech understanding, especially in groups of people, the interpretation of trackerless 3D human gestures and emotion and overall system responsiveness.

Potential Applications

Entity based

- Force on force combat simulations
  - "White Cell" controller to allow humans to give spatial hints to teams of "cognitively driven, perceptually based" human avatars
- Sensor layout studies
  - Improvement of the humans spatial perception of terrain impacts

Concept based

- Human perception based computer optimization
- Enterprise modeling
- Logistics modeling

- Multi dimensional data queries

References

Bowman, D. et al. Testbed Evaluation of Virtual Environment Interaction Techniques. Proceedings of ACM Symposium on Virtual Reality Software and Technology. 1999

Dybkjaer, L., Bernsen, N.O., and Minker, W. Evaluation and usability of multimodal spoken language dialogue systems. Speech Communication. 2004.

Mazalek, A., Davenport, G., and Ishii, H. Tangible Viewpoints: A physical approach to multimedia stories. ACM Multimedia. 2002.

Mcdonald, M. and Rigdon, JB, Umbra's System Representation. SAND2006-061506153819. 2005.

Oviatt, Sharon. Taming recognition errors with a multimodal interface. Communications of the ACM, vol. 43, No. 9, pp. 45-51. September 2000.

Paternò, F. Model-based tools for pervasive usability. Interacting with Computers. 2005.

Stanney, K.M., Mollaghasemi, M., Reeves, L., Breaux, R., and Graeber, D.A. Usability engineering of virtual environments (VEs): Identifying multiple criteria that drive effective VE system design. International Journal of Human-Computer Studies. 2003.

Ullmer, B. and Ishii, H. Emerging frameworks for tangible user interfaces. IBM Systems Journal, vol. 39, 3 & 4, 2000.

Ullmer, B. and Ishii, H. The metaDESK: Models and prototypes for tangible user interfaces. UIST Banf, Alberta, Canada. 1997.

Suggested Reading

Cohen, I and Li, H. Human Body Posture Inference for Immersive Interaction. ACM, International Workshop on Immersive Telepresence. 2002

Cohen, I and Li, H. Inference of Human Postures by Classification of 3D Human Body Shape IEEE International Workshop on Analysis and Modeling of Faces and Gestures. 2003

Cohen, I and Lee, MW. 3D Body Reconstruction for Immersive Interaction. Second International Workshop on Articulated Motion and Deformable Objects. 2002

D Rosenfeld, M Zawadzki, J Sudol, K Perlin. Physical objects as bidirectional user interface elements. IEEE Computer Graphics and Applications, 2004

Ishii, H. and Ullmer, B. Tangible bits: Towards seamless interfaces between people, bits, and atoms. Proceedings of CHI. March 22-27. 1997.

Kato H, Billinghurst M, Poupyrev I and Imamoto, K Virtual Object Manipulation on a Table-Top AR Environment. ISAR 2000

Kato, H. and Billinghurst, M. Marker Tracking and HMD Calibration for a video-based Augmented Reality Conferencing System. 2nd International Workshop on Augmented Reality. 1999

Kato, H. and Billinghurst, M. Marker Tracking and HMD Calibration for a video-based Augmented Reality. 2nd International Workshop on Augmented Reality (IWAR 99)

Lee, CM. and Narayanan S. Towards detecting emotions in spoken dialogs. IEEE Transactions on Speech and Audio Processing. 2005.

Patten J et al. Sensetable: a wireless object tracking platform for tangible user interfaces. CHI, 2001

Patten J, Ishii H and Pangaro G. Sensetable: a wireless object tracking platform for tangible user interfaces. CHI 2001

Potamianos, S. and Narayanan S. Adapative categorical understanding for spoken dialog systems. IEEE Trans. Speech and Audio Processing. 2005

Distribution:

2	MS 0899	Technical Library, 4535
1	1002	P. Heerman, 6630
1	1002	R. Skocypec, 6640
1	1002	S. Roehrig, 6600
1	1004	B. Hart, 6631
1	1004	D. Hart, 6631
1	1004	E. Hinman-Sweeney, 6631
1	1004	E. Parker, 6641
1	1004	F. Oppel, 6631
1	1004	G. Whitford, 6631
1	1004	K. Summers, 6631
1	1004	M. McDonald, 6631
5	1004	S. Tucker, 6641
1	1011	R. Cranwell, 6643
1	1188	B. Clark, 6641
1	1188	J. Wagner, 6641
1	1188	P. Buechler, 6641
1	1188	R. Abbott, 6641
1	9211	Central Technical Files, 8945-1