LINKING AUDIO AND VISUAL INFORMATION WHILE NAVIGATING IN A VIRTUAL REALITY KIOSK DISPLAY

BY

BRIANA M. SULLIVAN
Bachelor of Computer Science, University of Massachusetts, Lowell, 2002

THESIS

Submitted to the University of New Hampshire
In Partial Fulfillment of
The Requirements for the Degree of

Master of Science
In
Computer Science

May, 2004

This thesis has been examined and approved.

Thesis Director, Colin Ware,
Professor

Zachary Rubinstein,
Associate Professor

R. Daniel Bergeron,
Professor

Date

ACKNOWLEDGEMENTS

I would like to thank Colin Ware, my thesis advisor, for the great journey this thesis has been. His patience, support and trust encouraged me to do more than I ever would have imagined I was capable of.

I also want to thank Matthew Plumlee who began work on this project before I took over. His mentoring and constant kindness in dealing with the many questions I had allowed me to grow to a level of confidence that enabled me to successfully finish this research on time.

I thank my husband, Pete, who was by my side cheering me on for the entirety of my time doing this research. He helped me to feel balanced in the overwhelming tasks I was determining to undertake. Thank you Pete for everything.

I also need to thank my thesis committee, Dan Bergeron and Zack Rubinstein, for their time, criticism and advice throughout this process

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

LINKING AUDIO AND VISUAL INFORMATION WHILE NAVIGATING IN A VIRTUAL REALITY KIOSK DISPLAY

by

Briana M. Sullivan

University of New Hampshire, May, 2004

Computer based 3D museum exhibits that provide means for interaction with objects in a virtual world should to be easy to use, entertaining and informative. They mush enable users to quickly learn to navigate through a virtual world so the remainder of their visit can be spent on learning the educational content. This thesis is concerned with the problem of activating audio descriptions of images in such an exhibit while the user is navigating through it. By finding an easy-to-learn method of activating the audio descriptions, the cognitive load on users can be reduced, and their working memories freed up for the actual content of the exhibit. The main goal of this research is investigating various methods for activating audio descriptions in order to find the most effective. Effectiveness is defined as the method that inspires the user to spend more time viewing images with audio descriptions playing. The research has four major components: observing public use of existing exhibits to understand how they are generally used, designing and implementing an interactive 3D computer based exhibit, creation of different audio activation methods, and evaluation of these methods. The

results of this research indicate that active involvement in audio clip activation captures the users' attention better than more passive methods.

INTRODUCTION

Modern computer technology has made possible 3D interactive public kiosks that provide the user with a visually rich environment of multiple media that may include text, graphics, images, sound-clips, video, and animations. These environments also allow the user to interactively select content and navigate through the 3D environment to retrieve information. However, the navigation task may distract the user from the informational content. Ideally, the user should enjoy the benefits of the interactive 3D multimedia environment without sacrificing the ability to acquire the information it contains. Developing these types of interactive environments is a complex task due to the specific requirements of kiosks: i.e. they should be easy to use, as they must be proficiently operated within a few minutes; they should be self-explanatory as there are no human helpers; and they should engage users with interesting content so their experience is a memorable one.

This thesis is concerned with 3D interactive public kiosks and the particular problems of effectively linking visual 3D images with recorded spoken descriptions while a user is navigating. Multimedia, cognitive, and learning theories suggest that the cognitive load placed on users by aspects of the kiosk, which are not needed for learning the educational content, should be minimized [Schaller & Allison-Bunnel, 2003; MacDonald & Vince, 1994]. This requires finding an appropriate method for activating audio descriptions that is simple to learn and use.

We would like to understand how, for example, does a member of the public visiting a kiosk display make a connection between all of the items visually displayed and any sounds that may be associated with them? How can an interface be designed so users can make the connections easily, quickly, and naturally by themselves, without hindering their ability to navigate around the virtual world collecting bits of information about that world? There are numerous ways to both navigate and make connections between sounds and objects (selection) in a 3D environment [Hinckley et al., 1994]. Normally navigation and selection involve two separate modes but two modes can result in too much to learn in the short amount of time available to a kiosk visitor. Some navigation and selection techniques involve only a single mode. For example, activating audio corresponding to an object can be initiated in a number of ways: by avatar proximity to a particular object, by using the camera viewpoint, by the user's press of a button when near an object [Spaargaren et al., 2002], or even bumping the avatar into an object.

This research also has a practical goal. We had obtained a contract from the New Hampshire Seacoast Science Center to design and build the interface for a 3D kiosk intended to inform the public about aspects of the Marine environment. It was already established that this would be a stereoscopic computer display with a fly-thru interface and that the main content would consist of video and still images distributed through the 3D environment. The challenge, was to develop a technique enabling users to make audio-visual connections easily, quickly, and naturally by themselves, without hindering their ability to navigate around the virtual environment.

, which is to build a public kiosk for the Seacoast Science Center in Rye, New Hampshire. The results of this thesis will help to determine an appropriate method for activating audio clips that will be implemented in the kiosk at the science center.

This thesis is organized in the following way: Chapter 1 discusses the background theories and studies that contribute to a better understanding of multimedia in a 3D virtual kiosk type of environment. Chapter 2 discusses the specific requirements of the exhibit. The requirements include aspects that fit into the exhibit model given by the Seacoast Science Center, and time limits that help to determine how much time can be apportioned to learn the interface. Chapter 3 discusses five audio activation methods in detail. Chapter 4 introduces how the audio activation methods were evaluated. Chapter 5 presents concluding remarks on the work that has been done as well as directions for future work with audio activation methods.

CHAPTER 1

BACKGROUND

There are a number of areas of prior research relevant to issues dealing with 3D virtual kiosks. These include: the cognitive issues of how people learn; theories that have been developed to account for why multimedia presentations can be more effective; studies of how to control the users attention; studies relating to the best way of connecting images with audio while navigating; studies of whether active learning environments are better than passive learning environment; and virtual museum environments that have already been developed.

1.1 Cognitive Issues of How People Learn

Learning involves storing information in memory so that it can later be retrieved. There are numerous temporary demands placed on a user of a computer system that incorporates novel interfaces and environments (such as 3-D virtual worlds), which may make learning the interface and the content more difficult [Hitch, 1987]. Some of these demands require the user to remember temporary labels that might mark locations or parameters that delineate the boundary of the virtual world. The user may have a main goal to explore the virtual world but will also have to remember many sub-goals that lead to the accomplishment of the main goal, such as obtaining informational content at specific locations, and navigating to those locations. The user must also keep track of his/her current location within the virtual world along with what actions caused which responses by the system. Moreover, the user must remember the meaning of the current state of the computer; for example, if the computer is in an introduction mode then the user may not be allowed to navigate freely until the computer switches to the journey mode. All of these items must be remembered by the user in order to be effective in his/her interaction with the virtual world. With some insights from cognitive psychology it becomes clear that the mind can quickly be overburdened with too much information to remember.

Cognitive psychology is the study of the mind; it is concerned with the acquisition and use of knowledge, and with mental structures and processes. Central to modern cognitive theory is the concept of working memory. Working memory (a.k.a. short term memory) is a limited temporary store of information used during cognitive processes [Baddeley, 1986]. Abundant evidence shows that working memory is not a unitary structure but has separate components for visual information and verbal information. Baddeley's model of the architecture of memory divides working memory into a visuo-spatial sketchpad that works with 2 and 3-dimensional information, a phonological loop that deals with verbal information, and a central executive coordinating processor. Both the sketchpad and the phonological loop are specialized temporary memory systems, which are controlled by the central executive. The central executive is very active, being responsible for storing information regarding the current active goals, the intermediate results of cognitive processes, and expected inputs from sequential actions. It organizes (encodes) the information before it sends it to one of the "slave" systems in order to efficiently retrieve the information from memory when needed. Depending on the kind of information being processed (visual or verbal) determines where it is stored (in the sketch pad or the phonological loop, respectively).

Visual and verbal working memories support two mostly independent processing channels, one visual and one verbal. This is called dual-coding theory [Paivio, 1986; Clark 1991]. Similar to Baddeley's model there are differences in the way information is stored (coded) depending on whether it is verbal or visual. Verbal stimuli are processed through the auditory channel and the associated information from speech is passed to the verbal system for coding. Visual stimulus is processed through the visual channel and the information from any images is passed to the nonverbal system for coding. However, visual text is processed in the visual channel but coded in the verbal system. These models can help explain why remembering a picture with text is easier than remembering text alone. For example, if a student is shown a picture of a dog with the label "dog" below it the student will process the picture of the dog in the visual channel then stores it in the non-verbal system of working memory. However, the label "dog" will likewise be processed in the visual channel but then it will be passed into the verbal channel for encoding in the verbal system of working memory. An internal link will be connected to the picture of the dog and the label "dog" which will strengthen the relationship between them. When the student is asked about the "dog" s/he then has two places in working memory that contain information about the dog instead of just one. Therefore, a picture with words excites both the verbal and the visual processing systems whereas spoken (or written) words alone only excite the verbal system. It is believed that this dual excitement (or dual coding) is more effective than excitement of a single system. If learners can construct linked visual and verbal modals of mental representations, they learn the material better [Mayer & Sims, 1994; Mayer & Moreno 1998a].

These models agree on the major assumptions that short-term memory capacity is limited, and that information stored in short-term memory is quickly forgotten unless it is maintained by active control processes. For these reasons, public displays must not overburden working memory with information that is irrelevant to the educational purpose of the display.

1.2 Multimedia Theory

Mayer and Moreno propose that five active cognitive processes are involved in learning from multimedia presentations: selecting words, selecting images, organizing words, organizing images, and integrating words and images [Mayer & Moreno, 1998]. This has become known as the SOI (Select, Organize, and Integrate) model of meaningful learning. Selecting words and images equates to building mental representations in verbal and visual working memory (respectively). Organizing words and images consists of building internal connections among either the propositions or the images (in that order). Integrating implies building external connections between a proposition and its corresponding image. This can be shown with an example of using multimedia to teach how a bike pump works. The student is presented with pictures of a bicycle in an outdoor environment with a flat tire and a pump nearby. S/he hears a narration of the scene that talks about the components that make up a bike pump (handle, piston, cylinder, valve, air, hose, etc.) It is then the student's responsibility to select the words that correspond with the appropriate pictures, and then organize the words so they make sense with respect to the images seen. After the words and images are organized the student will begin to make connections between the words heard and their corresponding images. These connections lead to meaningful learning.

1.3 <u>Control of the Users Attention</u>

Directing attention of the user is an important aspect of a multimedia-learning environment because it allows the user to become conscious of what is to be learned. There are top-down and bottom-up factors that influence visual attention [Corbetta & Shulman, 2002]. The top-down features are cognitive (expectations or knowledge) while the bottom-up ones are sensory stimulations. Other factors that may influence attention are an interaction of the cognitive and the sensory resulting in unexpected or new experiences. A good interface design will contain elements that attract attention toward the most critical information resulting in active processing [Tellevik et al., 2003].

In human to human communications a common way that people link what they are saying to something in the local environment is through deictic gesture. Deixis is the act of drawing attention to an object or activity by means of a gesture. For example, someone points to an object and says, "Put *that*", and then pointing to another location says "*there*". Pointing denotes both the subject and the object of the command; verbal and visual objects are thus linked by deixis. In human-human communication speech and gestures are synchronized in time [Kranstedt et al., 2003]. There is evidence that deictic gestures, such as pointing, tend to occur at the beginning of an expression [Oviatt et al., 1997].

1.4 Linking Images With Audio While Navigating

"Deictic techniques can be used to bridge the gap between visual imagery and spoken language… A major advantage of combining gesture with visual media is that this multimodal communication results in fewer misunderstandings." [Ware, 2000, pg. 326] Deictic gestures/visual representations, with or without speech in a 3D world, assist the user in locating objects within that world.

Connecting images with audio through deixis while navigating is the function of some virtual pedagogical agents such as Cosmo [Johnson et al., 2000], whose purpose is to aid in the study of spatial deixis. Johnson et al. [2000] defines spatial deixis as "the ability of agents to dynamically combine gesture, locomotion, and speech to refer to objects in the environment while they deliver problem-solving advice." Cosmo has an internal planner that coordinates the agent's movements with its gestures and speech. Therefore, it can move towards an object, point at it and then speak about that object.

Mayer and Anderson [1992] have outlined, in their contiguity principle of multimedia, three forms of construction that the learner may make when connecting images with words: "(1) representational connections between verbal information that is presented and the learner's verbal representation of that information; (2) representational connections between pictorial information that is presented and the learner's visual representation of that information; and (3) referential connections between corresponding elements in the learner's verbal and visual representations." They also suggest that it is easier for students to make referential connections when images and words are presented contiguously due to the limited space in working memory.

1.5 <u>Active Learning vs. Passive Learning</u>

It is widely held that it is the activity of making connections between images and words that aids in learning. As a Chinese proverb says (attributed to Confucius):

Tell me, I forget.

Show me, I remember.

Involve me, I understand.

The assumption that active learning is important in the learning process is the foundation of a constructivist approach to learning [Duffy & Cunningham, 1996]. The dominant theory of learning applied to education is the constructivist theory that learning is an active cognitive constructive process where the learner interacts with the world, and builds a meaningful mental representation of external reality [Mayer et al., 1999]. Constructivism is the idea that knowledge is not simply acquired but is constructed by individuals based on their own personal experiences. The need for active involvement is also central to situated learning theory [Lave & Wenger, 1991; Brown et al., 1996], and engagement theory [Kearsley & Shneiderman, 1998].

Researchers have studied the benefits of using Virtual Reality (VR) from a constructivist perspective, and claim that VR provides "direct and intuitive interaction with information" [Bricken & Byrnes, 1993], thus stressing the importance of using VR within a constructivist framework. Other findings also support the idea that 3D virtual worlds can be used as learning environments [Bricken & Byrnes, 1993; Kelly, 1993].

Activity theory is intended to provide a framework for building constructivist-learning environments [Jonassen & Rohrer-Murphy, 1999]. It proposes that activity is a precursor to learning instead of traditional views of learning before doing. As we act, we gain knowledge, as we gain knowledge we adjust our actions, which in turn changes our knowledge, and so on. In other words, there is mutual feedback between knowledge and activity [Fishbein et al., 1990]. Interaction therefore, is assumed to help the learner recall the information that was learned [Dick & Cary, 1990], and thereby plays an important role in the process of understanding [Brooks, 1988; Hibbard & Santek, 1989].

Empirical studies involving children have shown that active exploration of an environment results in a better understanding of spatial relationships than do purely passive experiences [Hazen, 1982; Feldman & Acredolo, 1979; Cohen & Weatherford, 1981]. The main conclusion is that active choice is the significant factor in good performance in a search task.

Despite studies showing that active learning can be valuable, there has also been research suggesting that active experiences in an immersive virtual environment results in no difference in learning than passive experiences in the same environment [Wilson et al., 1997; Wilson, 1999; Melanson et al., 2001].

A major concern related to the desirability of active exploration is that in a complex 3D environment with input controls and system behaviors that are novel, attention could be taken away from learning the content of the kiosk and redirected to simply learning the controls and responses of the environment. This might negate the positive benefits of active exploration.

1.6 Virtual Museum Environments

There have been many forms of virtual museum environments developed [Gaitatzes et al., 2001], most based on the constructivist model of learning. Many are web-based and include platforms that allow tours of one form or another. Some have implemented a “virtual tour” as a simple 360-degree panoramic view of a given location, while others allow movement through different rooms of a museum. These systems differ in the particular ways in which they allow the user to navigate through the virtual space and how they attempt to make the connection between words and the visual displays of interest.

1.6.1 Navigation

Navigation in computer environments can take many forms. People familiar with the World Wide Web know about hyperlinks used for navigation. Using a hyperlink is a simple as clicking on a word, phrase, or picture that instantly triggers a location change. There are other forms of navigation associated with computer games that are also familiar to everyday types of transportation: driving, walking, and flying to name a few. These metaphors have also been adopted in the virtual environments of the World Wide Web and interactive museum kiosks. The controls for these types of metaphors can vary from keys on the keyboard, a mouse input device, a joystick or even a steering wheel. Following are examples chosen to illustrate different navigation methods.

The interactive *Explore the Fort at Mashantucket* [Hanke, 1992] provides a 3D re-creation of a 17th-century Pequot fort. This exhibit takes visitors on a guided tour of this ancient site with the help of 3D computer animation. On-line visitors can "navigate" through the fort by clicking on arrows. As they are clicked, documentary video, and artifacts with associated text descriptions pop up in a smaller window above the layout view of the village. This form of navigation is based on a hyperlink method; the actual exhibit on-site allows a full CGI motion version of the 3D video animation.

The Boats Afloat exhibit was the winner of Massachusetts Interactive Media Council (MIMC) Award, 1996, Best Kiosk [Larson, 1996] for its novel interface design. It is a state of the art interactive video that allows museum visitors to "drive" six different boats through the Boston Harbor by moving a real boat throttle and steering wheel. The throttle and steering wheel are mounted on a twenty-foot lobster boat. Visitors can watch a monitor that displays their location and speed on a map. Rumbling boat sounds that come from subwoofer speakers indicate the changing boat speed.

The Virtual Museum of Computer Science in Italy [Barbieri et al., 2001] is a Collaborative 3D Virtual Museum of Computer Science. It can be visited online, as a traditional web site, but it also provides a three-dimensional virtual reality interface set in a science fiction type of setting. It can allow people from different parts of the Internet to visit it together. It also contains simple interactive games that illustrate basic principles of computer science. The visitors choose an avatar to represent them, which can navigate through the 3D world by flying or jumping from various floating platforms in the sky. The jumping is accomplished by clicking on the desired destination location in view; as this is done, the avatar is relocated in a smooth animated fashion as if the avatar is "jumping". There is another mode that allows navigation to be done by having the avatar follow the cursor (where the mouse is positioned in the 3D world), which results in the appearance of the avatar "floating" through the 3D world. To get help a user can click informational billboards, which display text or web page links. A virtual tour guide can lead users through the virtual world to help them discover games and explains, teaches and discusses topics with the user as it leads the user to various locations. Communication with others or a virtual tour guide occurs through the keyboard in a chat-room type of fashion.

1.6.2 Augmented Reality

An important question regarding modern uses of technology in museums and virtual environment is how can users access audio narrations that are linked to the images they describe? In many example web sites it is common to click on objects with the mouse button to play associated sound. Some exhibits within actual museums have more innovative methods of interaction that are unique alternatives.

The exhibit at Castello del Buonconsiglio in Trento is actually an interaction between the visitor and the museum itself. This type of interaction is possible by informing the underlying information system the physical position of the visitor (and as much as possible his visual focus). As the visitor moves around the museum, the system communicates, via portable devices (such as a palm pilot, headphones, or head-mounted display) 3D audio and/or superimposes generated images (via projectors) onto the real scene. The system receives requests from the visitor verbally or through gestures [Stock & Zancanaro, 2002].

There are museums where the visitor can use a wireless headset to obtain information about the objects within the museum. Guide-Man, created by Ophrys Systems in France [Guide-Man, 2002], is an audio-guide system (similar to a mobile phone) used mainly for museums and cultural sites. It provides different methods for commentary on exhibits or points of interest. Some of the methods include: the

Transponder method (activation of commentary from a specified short distance), the Remote-Control method (activation of commentary by a tour guide), Radio-Frequency method, (activation at a long distance – from 1m to 100m, and able to be synchronized with a video display located at the exhibit/point of interest), and an Infrared method (activation within a specific zone and synchronization with video). That means, for example, a user can hear the object information as s/he approaches the object of interest, or s/he can push a button when s/he is in front of the object of interest. There is also a mode in which a human tour guide with a group of visitors directs them to the next item of interest and s/he pushes a button to activate the commentary once they are in position to view the object. The Guide-Man also includes user control over rewinding, playing or pausing the audio.

In a VRML (Virtual Reality Markup Language) system developed for people with disabilities, Ressler and Wang [1998] provided three different methods that define when a spoken description/sound is to play: proximity (execution based on viewer position), viewpoint (execute when selecting viewpoint), and touch (execute on viewer clicks). An example of the proximity method was used in an assembly line world. As users approached specific objects in the assembly line they would enter a 3-dimensional box. Upon entry into the box (unseen by the user), audio associated to the object being surrounded would begin to play. They also associate sound nodes with spoken descriptions of objects of interest. However, they did not carry out research to determine the best method to use in a given situation.

CHAPTER 2

EXHIBIT REQUIREMENTS AND DESIGN

2.1 Seacoast Science Center Exhibit

A major motivation for this study was the need to develop a functioning exhibit at the Seacoast Science Center. The Seacoast Science Center is a small museum located on the coast of Rye, New Hampshire devoted to providing information about ocean activity. This provided both the opportunity for the research and imposed a number of constraints.

The exhibit/public kiosk (GeoZui3D Explorer 1) that will be used as the platform for our tests has been built using GeoZui3D [Ware et al., 2001] as a starting point. GeoZui3D is a highly interactive 3D visualization system designed to assist in the research of georeferenced data (Geo), which has a Zooming User Interface (Zui), developed by Roland Arsenault, Matthew Plumlee, Andy Foulks, Don House, Larry Mayer, Shep Smith, and Colin Ware in the Data Visualization Lab (part of the Center for Coastal and Ocean Mapping) at the University of New Hampshire (UNH). The preliminary work on GeoZui3D Explorer 1 evolved as a joint effort between Tracy Fredericks, the Special Projects Director of the Seacoast Science Center (in Rye, New Hampshire - which partially funded it), Matthew Plumlee a PhD student of Computer Science from the Data Visualization lab, and Colin Ware. I took the lead in the development beginning in June 2003.

The exhibit is set up in a section of the Seacoast Science Center that has similar themes in the area. The computer, stereo display screen, shutter glasses and input controller are encased in a large wooden kiosk structure as shown in (Figure 1).

Figure 1. GeoZui3D Explorer 1 kiosk at the Seacoast Science Center.

GeoZui3D Explorer 1 is capable of displaying the following objects:

- A digital terrain model of the seabed (Bathymetry)
- Movie clips shown as billboards in the 3D environment. These can have text labels
- Images also shown as billboards with text labels

- A 3D Dive sled that acts as a proxy for the user. This is placed ahead of the user, and is driven around using a control yoke.
- Other 3D models. For example, a 3D model of a bridge can be placed in the scene.

The bathymetry is a 3D data representation of the bottom of the Piscataqua River (approximately 800 meters across). A color-coded chart, which is displayed to the lower left of the users screen, represents different levels of depth in the river (Figure 2).

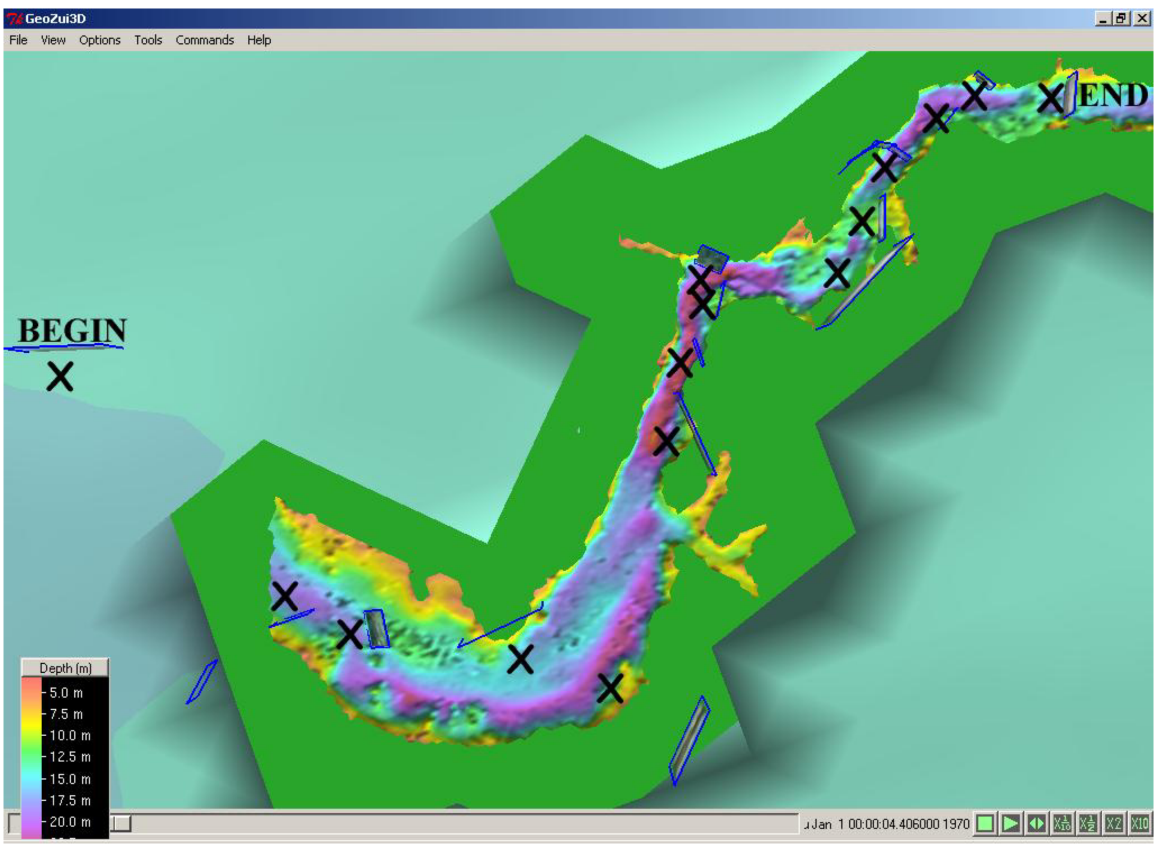

Figure 2. The overhead view of the kiosk virtual environment. It shows the Piscataqua River and the location of the points of interest.

Throughout the length of the river are various locations where movie clips or images are situated. Each image/movie clip is labeled with a title below it. Also associated with each image/movie clip is an audio clip that gives any where from 5 to 28 seconds of information about the corresponding image/movie clip. An overview of the layout of the virtual environment is shown in Figure 2, with an “x” at each location where there is an image with an associated sound clip. There are a total of 14 images with associated sound clips that the user can freely navigate to.

The boundary of this virtual journey is delineated by the digital terrain model. If the user drives into the edge or bottom of the virtual river, the vehicle is pushed back the way it came. The force on the vehicle backwards is comparable to the speed with which the vehicle intercepted the boundary. There is also an upward cap on how high the vehicle can rise. This is a set limit around the level of the highest part of the river channel. This upward limit helps give the feeling of being “inside” the river.

The “dive sled” is the vehicle that the user uses to explore the river (represented as approximately 10 meters across). This acts as a kind of proxy (or avatar) for the user within the scene, although it is placed ahead of the user to keep it in the field of view (see Figure 3).

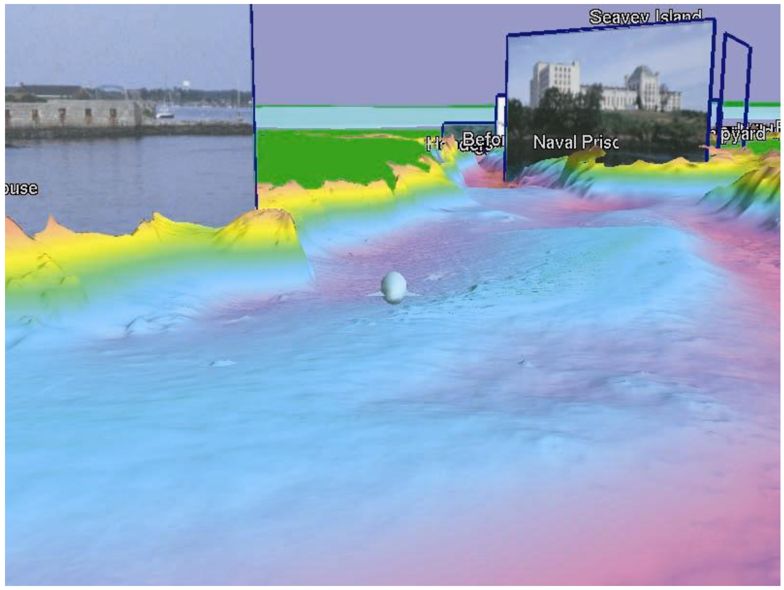

Figure 3. GeoZui3D Explorer 1 – User view journeying up the Piscataqua River (the dive sled is located in the center of the figure).

The vehicle is maneuvered using the input controls illustrated in Figure 4. This is a rugged device, designed for use by the arcade video game industry.

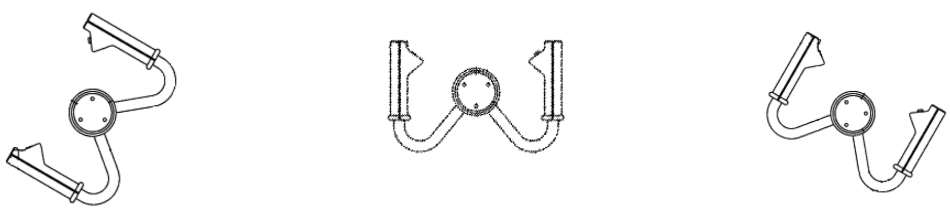

Figure 4. Navigational device used for exhibit input. Left, middle and right positions.

The forward velocity is controlled by pushing the entire device forward. The rate of turn of the vehicle is controlled by the displacement of the rotation of the device (clockwise or counter-clockwise). There are two trigger buttons at the front of the steering device (one for each hand) that are used for menu selections, and possibly "shooting" images in the tour to activate their corresponding audio descriptions.

The user arriving at the kiosk first encounters an introduction screen with a label across the bottom of the screen that says, "Press a trigger button to begin". In the background is an animated introduction screen that shows an overhead view of the coastline from Boston, MA to Maine. There are labels showing where major sites (i.e. Boston, Portsmouth, the Piscataqua River, and the Seacoast Science Center) are located. The user does not have control at this point; instead the users' view is animated by the system taking them from an overview of the Gulf of Main to a "landing" in front of the Seacoast Science Center. During this time, an associated audio clip is played

> *"Welcome, to GeoZui3D Explorer 1. This exhibit lets you choose a virtual journey up the Piscataqua River, to an aquaculture site near the Isles of Shoals, or to George's Bank. Along the way you will encounter video clips and 3D images of sea life and other interesting attractions situated over the latest bathymetric model of the seabed."*

As the viewpoint approaches the Seacoast Science Center image the user hears, *"You are now flying down into a 3D computer model of the Gulf of Maine."* The other journeys, Isles of Shoals and George's Bank, are to be completed in the future, and will have an associated menu allowing the user to choose between the different journeys. Currently the Piscataqua River journey is the default journey.

When the user presses a trigger button to begin the virtual experience a menu appears in the middle of the screen that has two choices (Figure 5). A recorded explanation then begins to play, "You have two choices, you can take a self-guided tour where you will control the vehicle yourself, or you can let us take control and guide you.

To make a choice, turn the controls in the direction of the type of tour you would like, then hold it there while you push a trigger button."

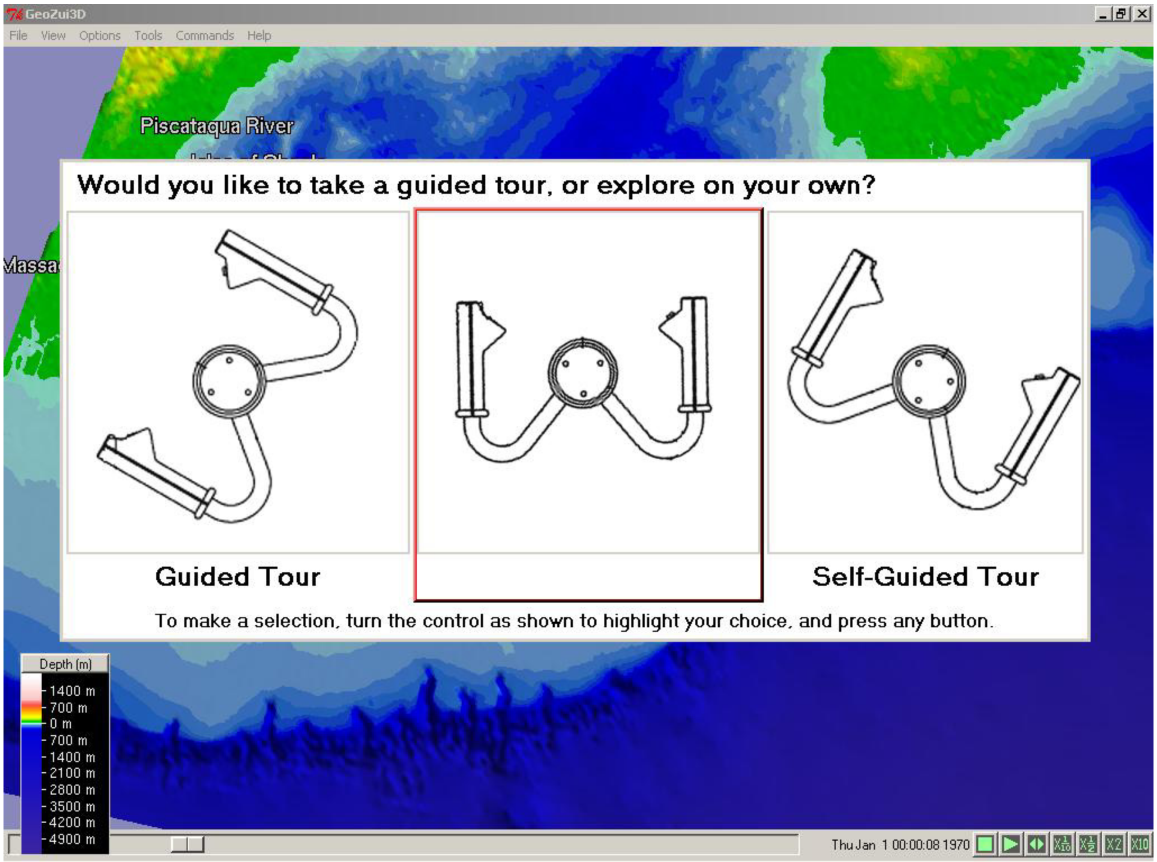

Figure 5. Menu selection at the start of the virtual journey.

The user then sees the dive sled position itself in front of an image of the Seacoast Science Center with this recorded introduction, *"Your journey begins here at the Seacoast Science Center. You will travel up the Piscataqua River which ends at Great Bay."* The user is then transported by the system from the start, the Seacoast Science Center, to the beginning of the bathymetry near the first "x" in Figure 2 (in the upper left hand section of the figure). During the automated transportation of the dive sled the user hears a brief introduction to the bathymetry ("The bright colors you are seeing in front of you represent the bathymetry, or the different levels of depth in the river, which come from the state of the art multi-beam sonar survey") and the following directions on how to maneuver the dive sled, "To exit this journey at any time, press the button by your right thumb. To navigate the dive sled push forward on the handles to move forward, you can also tilt the handles to move up and down. Enjoy your journey." The introductory script is timed to end when the automated transport of the vehicle ends inside the bathymetry near the first image on the journey. At this point the input device is activated and the user is then free to explore the Piscataqua River at his/her leisure (Figure 3).

An iterative design approach was taken with this exhibit before any evaluative testing was done on the audio activation methods. The focus of this preliminary work is to ensure the user's ability to activate the audio descriptions of each image while navigating. To do this it was important to simplify the interface so it was very easy to use and took very little time to learn.

2.2 Assessment Of Length Of Stay At Exhibits

Because time is a major concern of kiosks, time requirements were needed to guide the development of the interface to be used. In order to obtain an understanding of the time requirements of a museum exhibit, a program of observation of various exhibits was undertaken at the Boston Museum Of Science (See Appendix A: Data from Length of Stay at Exhibits Assessment). Exhibits were chosen to represent a variety of different types, ranging from highly interactive/creative exhibits to observational exhibits with no specific objective. These were observed to gather information about the length of time visitors might typically spend engaged at such exhibits.

The result of this study showed that the least (average) amount of time spent on an exhibit was 32 seconds and the longest (average) amount of time spent on an exhibit was 9 minutes 18 seconds. In order to determine the length of stay a typical visitor would spend at the exhibit for the Seacoast Science Center we took the average obtained from the Virtual Fish Tank (see Appendix A.5). This exhibit was a highly creative computer-based exhibit that was most comparable to our exhibit Based on these observations we designed an interface for an exhibit with the expectation that the typical visit time would be around 5 minutes. With a 5-minute average time on an exhibit, we feel it would be reasonable to learn the novel interface within 1 minute.

CHAPTER 3

AUDIO ACTIVATION METHODS

The ideal audio activation method would encourage interaction with objects in the exhibit, so that verbal information associated with visual objects can be conveyed. To investigate the best audio activation methods, five techniques for playing audio were developed through an iterative design process carried out in conjunction with the Seacoast Science Center. These were influenced by methods described in the research literature, educational theory, and an interest in investigating the difference between active and passive modes of interaction. The broad goal in designing each audio activation methods was to have a method that encourages users to spend the majority of their time facing activated informational points within the exhibit.

The five methods can be visualized as a pyramid (Figure 6), each successive method building upon the characteristics of the one below it.

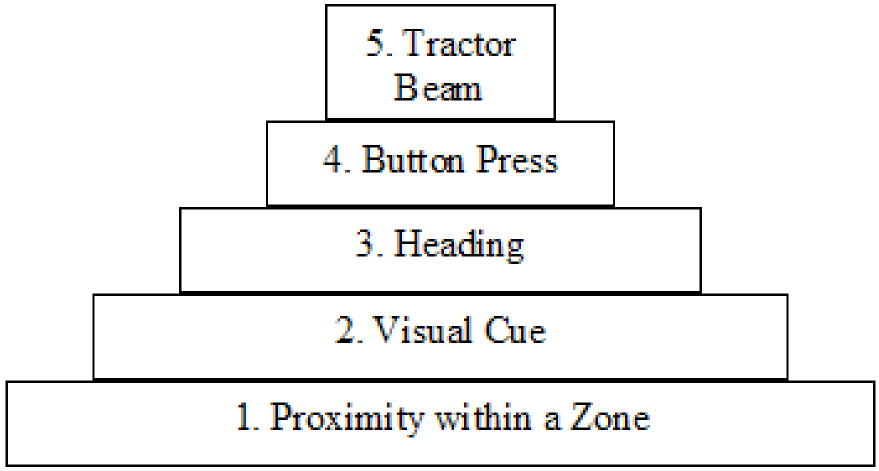

Figure 6. Pyramid of five methods to activate sound corresponding to an image.

There are trade-offs between the simplicity of the lower levels of the pyramid and the higher interactive upper levels of the pyramid. With a simpler interface, less cognitive resources are used to learn how to navigate and activate audio-clips than with highly interactive, more complex methods of navigation and activation of audio-clips. More interactive interfaces, higher on the pyramid, could deter non-computer game players from trying them out, while simpler interfaces could bore computer game players and deter them from continuing to use it. Prior to the study it wasn't clear which of the five would be the best due to these trade-offs.

3.1 Proximity within a Zone

The first level of the pyramid is labeled "Proximity within a Zone" and depicts the simplest method of activating sound for any given image of interest in the virtual journey. In the "Proximity within a Zone" method the vehicle triggers the audio attached to an image when it enters a specific zone defined as an area in front of the associated image (that has a 60 degree angle out from the center of the image). This is depicted in Figure 7.

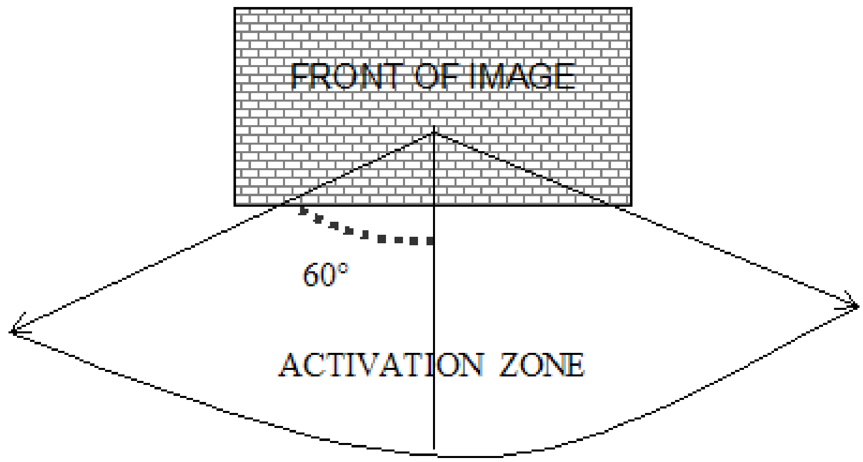

Figure 7. Activation Zone that plays audio clip associated with the image.

The size of the zone is proportional to the size of the image since smaller images will be seen better at a closer distance and larger images from farther away. This method of audio clip activation was chosen to be the easiest for the user to use since the only requirement is to drive the vehicle close to the image. This made it possible for us to test a method that allowed anyone who was able to drive the vehicle to access the audio for a specific image.

An anticipated problem with this design was that the activation zone would not be visually defined to the user. Because of this, the user might trigger an audio clip with several images in the field of view possibly causing a cognitive association between the audio clip and the wrong image.

3.2 Visual Cue

The second level in the audio clip activation methods pyramid is that of "visual cue" that tells the user when s/he has entered an audio activation zone. The visual cue takes the form of a blue line from the center of the vehicle to the center of the image within the activation zone. The blue line appears as soon as the vehicle has entered the activation zone and disappears the moment it leaves the activation zone. The blue line visual cue is synchronized with the playing of the audio clip associated with the image. The visual cue was added in the hope that it would lead the user's attention to the appropriate image corresponding to the audio clip. Even in the event that a picture is not in the visual field when the audio is activated (i.e. when the user is in the proximity zone, the visual cue will at least let him/her know which way to go to see the associated image.)

Again, the activation zone is still not clearly defined to the user. Although visual cues may help with this problem, it is still possible for an audio clip to start when the image is not in the current field of view.

3.3 Heading

The third level in the pyramid is that of "Heading". The heading device attempts to take care of the problem of the audio being triggered before the user can see the attached image. It is modeled after social behaviors of communication, namely that a person being approached to communicate with is faced in order to initiate communication. Heading is defined as the yaw of the vehicle or the side-to-side position of the front of the vehicle. A view angle of 35 degrees is used to determine if the center of the image is within the vehicle's heading zone, which causes the heading trigger to be activated (the vehicle must also remain inside the activation zone). This is shown in Figure 8.

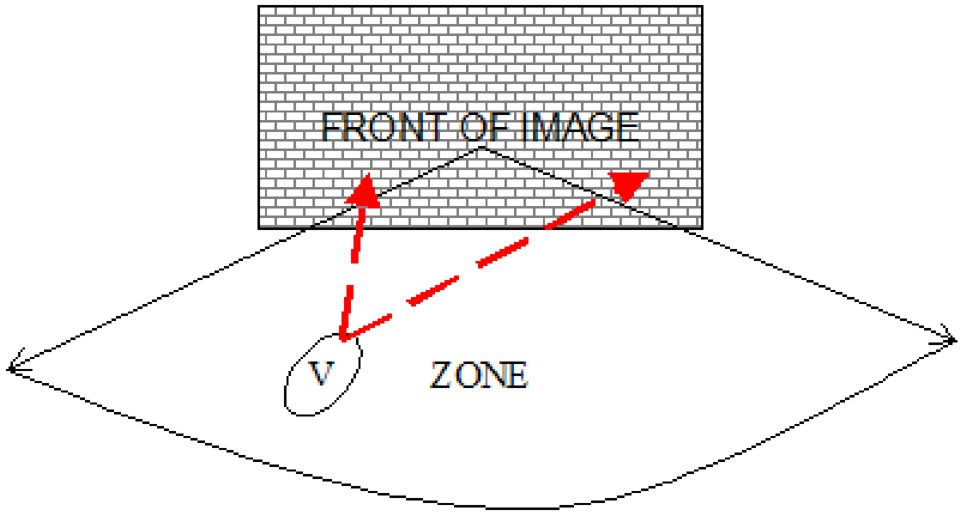

Figure 8. Heading method of activating sound associated with the image.

The blue line zone/audio cue becomes visible the moment the vehicle's heading is within the view angle and disappears the moment the vehicle leaves the zone or is no longer within the view angle.

All three methods discussed so far are passive; in the sense that no extra actions beyond navigation to a particular position is needed to activate audio. Constructivist theories suggest that explicit user action such as a button press would be desirable. This would allow users to actively control when an audio clip begins.

3.4 <u>Button Press</u>

The fourth level in the pyramid is that of "button press". This is a combination of all the previous levels, namely; proximity, visual cue, heading and a button press to activate the audio clip. The button press was added to evaluate the value of more active involvement in the activation of each sound clip. When the vehicle is within the zone and facing the image of interest, a thick highlighted yellow frame appears around the image (normally all images are outlined in blue). At that point, the user can turn the audio on by pressing a trigger button. Once the audio begins to play the blue line visual cue will appear then disappear when the vehicle has left the activation zone.

This method required additional instruction at the start of the exhibit. Users had to be taught to use the buttons as well as navigate with the steering device. In addition, there could still be some confusion regarding where exactly each activation zone was located, thereby causing the user to repeatedly press a button though they may not be in the correct location for the audio clip to be activated.

3.5 Tractor Beam

The fifth, or top, level of the pyramid is that of "tractor beam". This uses all of the components including the button press, but delays playing the audio until a "tractor beam" moves the vehicle to an "optimum" viewing position. The tractor beam is engaged when the button is pressed within the activation zone and when the vehicle is facing the image (again the yellow highlighted frame visual cue will appear). After the button is pressed, the user temporarily loses control of the vehicle and the vehicle is smoothly translated and rotated to the optimum viewing position. This optimum viewing position is defined as a position that is centered on the image and at a distance from the image where the image fits into the user's frame of view. When the vehicle has arrived at the optimum viewing location, the sound clip is activated (the blue line visual cue will also appear on audio activation then disappear when the vehicle leaves the activation zone). Control of the vehicle is returned back to the user after the tractor beam animation is complete enabling the user to continue navigation at any time.

Anticipated problems included possible confusion with loss of control as the tractor beam activated.

There are some issues with the audio activation zones and their close proximity to one another that applied to all of the audio activation methods. First of all, there are only two things that terminate the audio descriptions. One being the audio clip expires on its own, and the other involves activating another audio clip. Since the blue line visual cue (for all but the proximity audio activation method) appears when the audio begins, it is also possible for the blue line to appear for two different pictures if their audio activation zones happen to overlap. However, in that case the audio for the first image will stop with the new image activation zone overriding the audio with its' associated audio clip (and also a blue line will appear).

CHAPTER 4

EVALUATION OF AUDIO ACTIVATION METHODS

The five audio activation methods presented in the previous chapter were each intended as a plausible best solution to the problem of linking verbal explanation with imagery in a 3D navigational environment. These methods can be characterized on a continuum from least to most interactive. In particular the last two methods required active button presses to activate the audio material. Constructivist theory suggests that more active participation is beneficial; on the other hand, this makes the interface more complex and places additional burden on the user.

One of the goals of this thesis was to assess the value of constructivist theory when applied to navigation and audio activation in a 3D public kiosk. The aim is to answer the following questions; "Should users actively trigger speech events by selecting objects in some way or are more passive activation modes better?"

The measure chosen to assess the effectiveness of the different audio activation methods was the length of time users spend facing images while the audio is being played. We predicted that the more active methods of audio clip activation involving a button press would result in the users spending more time at informational points, listening to the audio content. This research was also intended to have a direct practical benefit; the best method will be used in the Seacoast Science Center exhibit GeoZui3D Explorer 1 irrespective of whether or not it supported the theories.

The evaluation of audio activation methods had two forms, objective (study 1) – using measured behavior with exit interviews involving the museum visitors, and a subjective comparative assessment (study 2) – using both observation and semi-structured interviews with volunteers who were exposed to all five audio activation methods.

4.1 Study 1: Testing with Museum Visitors

The evaluation of the five methods for activating audio sound clips was carried out with the general public in a museum setting. Unlike most human factors evaluations where people are told they are in an experiment, this was done in a natural museum environment. The visitors were told only upon request, that the exhibit was being evaluated (with no mention of what was actually being tested, so as not to influence their natural behavior)[*].

4.1.1 Subjects

The kiosk subjects were visitors of the Seacoast Science Center; when it closed for renovations, testing was moved to the New England Aquarium. For the objective tests statistics were collected from 100 subjects in total. 20 different users tried each of the 5 different audio clip activations methods. The breakdown of the subjects is shown in Table 1.

[*] IRB approved, See Appendix B.

Table 1. Breakdown of quantities of subjects in each category (method type, age: adult/child, gender: female/male)

T1				T2				T3				T4				T5			
A		C		A		C		A		C		A		C		A		C	
10		10		7		13		11		9		11		9		6		14	
F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M	F	M
5	5	5	5	3	4	5	8	5	6	1	8	4	7	4	5	2	4	2	12

The adult category was comprised of high school aged persons and older. During formal experimentation it was observed that younger children below 4th grade age do not seem to have the capability to be able to do more than one thing at a time (accelerate and turn for example), therefore this exhibit was not tested with this age group.

4.1.2 <u>Procedure</u>

As visitors of the science center approached the exhibit, they were informed (if asked) that the exhibit was being run on a trial basis with the general public. (At the Seacoast Science Center there were signs up saying “Exhibit Under Construction” on one of the two exhibit controllers, so it was necessary to let the patrons know that one side of the exhibit was open and available for use.) At the New England Aquarium a more portable exhibit was set up near the entrance of the aquarium. Since it was new and didn’t appear as a normal exhibit, it was also necessary to inform the patrons that it was indeed available for use.

The program for the virtual journey of the Piscataqua River was the same for all visitors (regardless of the location it was being tested in); however, for visitors that tested the button press and the tractor beam audio activation methods, it was necessary to explain to them during the automated intro to the journey (as described in Chapter 2) that the trigger buttons were to be used for "shooting" the pictures to obtain the audio information from them. This was necessary because the audio activation method being tested cycled about every 3-5 patrons (depending on the crowds) to ensure all methods could be tested on the same day with a similar crowd. This rendered it difficult to quickly change the prerecorded audio description explaining what the trigger buttons were for in the last two methods.

All subjects were observed as they were exploring the 3D virtual environment with no imposed time requirements. When each subject was finished with their exploration, they were approached and asked to participate in a brief exit interview. The exit interview asked questions about various parts of the exhibit. Three in particular were specifically designed to get at the issue of knowing how to activate audio and the effectiveness of the visual cues; the other questions related to user interface and preferences and are not discussed here. The three relevant questions were:

1. Did you know how to activate the audio description associated with each image in the exhibit?
2. Did you know what the blue line (visual cue) coming from the vehicle was for?
3. Did you notice the yellow frame highlight around the image (if using a button method)?

An exit interview was attempted with all participants; roughly three-quarters of the participants answered at least one of the questions posed. Various factors came into play as to why questions were not asked or why participants did not answer: some were unwilling, other museum visitors demanded attention, etc.

4.1.3 Measures

Specific data was collected internally through the program that gathered information such as the length of time for the overall visit, the time spent inside each zone, the time facing each picture, the overall count of images visited and the number of audio descriptions activated.

The overall time spent facing each activated informational point is used to estimate the level of interest a user has in that topic [Stock, 2002]. Two measures are used to assess the degree to which users access the content of the exhibit: a count of the number of activation zones entered by the user, and a measure of the length of time users spend oriented towards the images. The total number of zones entered is taken to indicate the ease with which a user could navigate through the environment. The length of time a user lingers in front of an image oriented toward it is used as a measure of the users level of interest in that image.

4.2 Results of Study 1

The results of this study are discussed in the following order; empirical measures, exit interview results, and observations made by the experimenter.

4.2.1 Empirical Measures

Two separate analyses of variance were run on the following factors: time facing audio activated images, and the number of zones entered.

The average time a user was oriented toward an image in a zone with audio activated was used as a measure of the users level of interest. This captured the fact that the user was probably both "looking at" an image and "listening to" the audio associated with that image. Using this measure as the dependant variable there were two main effects: the audio activation method used [$F(4,80) = 6.84$, $p < .0001$] and age [$F(1,80) = 7.52$, $p < .009$]. Tukey's post hoc Honestly Significant Differences (HSD) comparisons revealed 2 groups. The tractor beam (T5) and button press (T4) methods made up the first group (active methods) and showed average times of 9.8 seconds and 6.6 seconds respectively. In the second group were the zone (T1), visual cue (T2) and heading (T3) methods (passive methods) and had average times of 2.1, 2.0 and 2.6 seconds respectively. In addition there was an age-method interaction [$F(4,80) = 2.52$, $p < .05$] (summarized in Figure 9), a gender-method interaction [$F(4,80) = 5.68$, $p < .0001$] (summarized in Figure 10).

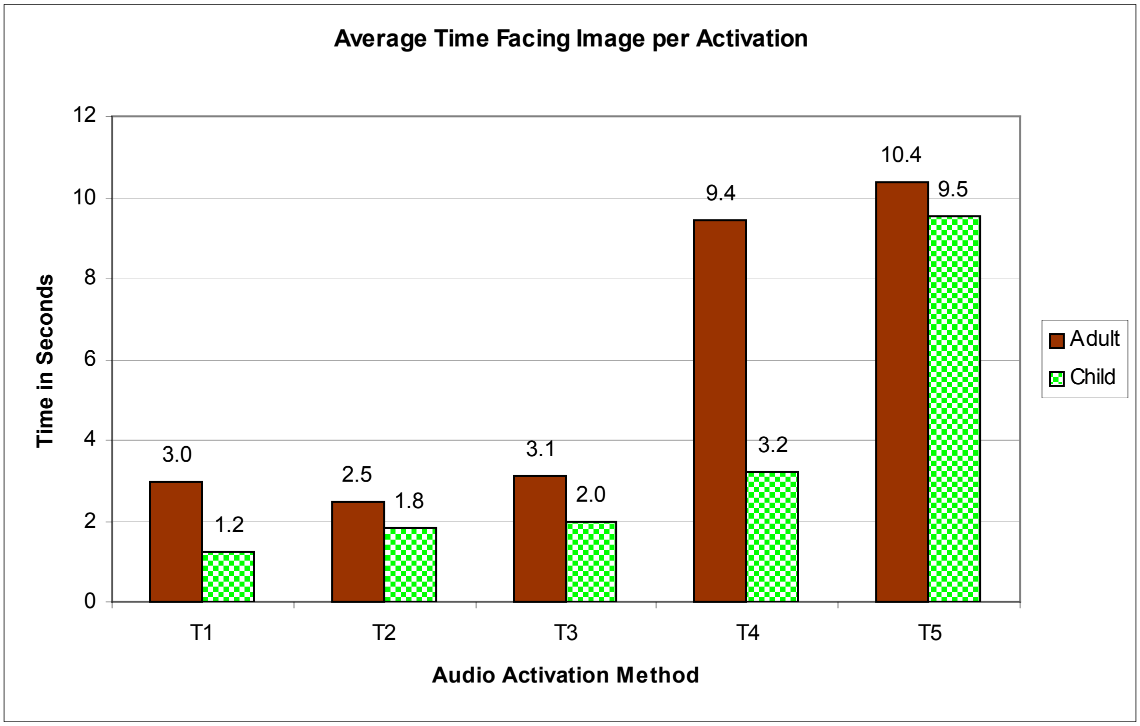

Figure 9. Average time facing images per activated audio clip according to age per method used.

Figure 9 shows that both adults and children face images (with audio playing) much longer when the tractor beam (T5) method is used (10.4 and 9.5 seconds on average, respectively). Adults spend more time than children facing images when using the button press (T4) method (9.4 sec).

The gender-method interaction (Fig. 10) shows that males face audio activated images longer while using the tractor beam (T5) method (11.5 sec), where females show greater times for the button press (T4) method (9.8 sec).

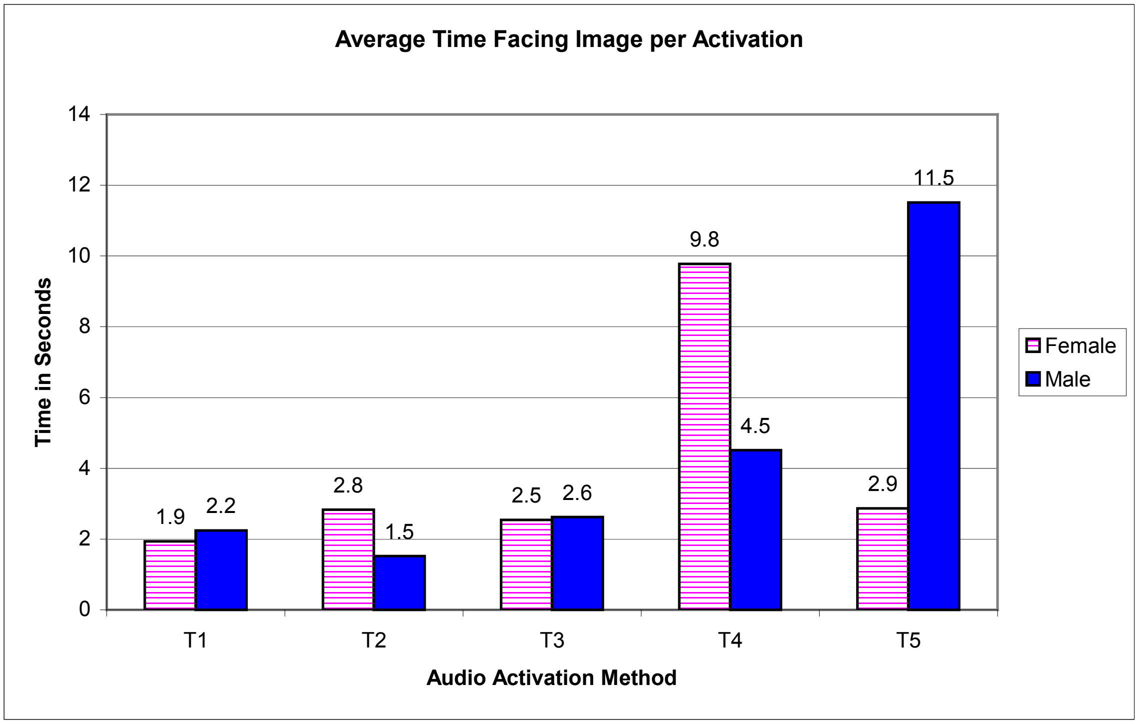

Figure 10. Average time facing images per activated audio clip according to gender per method used

Using the number of zones entered as the dependant variable the main effects were: the audio activation method used [$F(4,80) = 3.06$, $p < .03$] and gender [$F(1,80) = 7.77$, $p < .006$]. The average number of zones entered according to method are broken down as follows: zone (T1) = 7.5, visual cue (T2) = 8.2, heading (T3) = 5.1, button press (T4) = 5.3, and tractor beam (T5) = 6.2. In addition there was an age-method interaction [$F(4,80) = 2.82$, $p < .03$].

4.2.2 <u>Discussion of Empirical Results</u>

The main result was that users spent longer times in front of the images with the two active methods of activating audio (the button press – T4 and the tractor beam – T5). This suggests that these more highly interactive methods for linking images and sound produce a higher level of interest in the content presented than do more passive methods. Also, adults on average spent more time facing images than did children.

The age-method interaction indicates that adults and children are affected differently according to the method of audio activation they are using. Adults, using the button press (T4) method, spent three times as much time facing images with activated audio as children.

In the gender-method interaction, females had longer times facing images when using the button press (T4) method and males had longer times using the tractor beam (T5) method. This was an unexpected result due to the nature of the tractor beam. The tractor beam takes control away from the user when the user pushes the trigger button at the appropriate location. This was foreseen to confuse the user. It was believed that the user, once feeling a loss of control, would not know when exactly control would be returned. This would, in essence, "trick" the user into not trying to move the vehicle until the audio clip stopped playing. One possible explanation for these results is that males weren't aware of the return of control so they lingered longer.

The results of the analysis of variance using the number of zones entered as the dependant variable showed that there was no significant difference in the methods for the effect they had on the users ability to navigate through the environment. This suggests that the users ability to navigate was not affected by the audio activation method.

4.2.3 Exit Interview Results

The data recorded for the exit interviews is incomplete due to the time schedules of the people who were observed at the exhibit. Not all of the people involved in this study answered all of the questions presented; in some instances it was not possible to ask the user any questions at all.

Table 2 shows the “yes” answers to the three relevant questions (1. “Did you know how to activate the audio description associated with each image in the exhibit?”, 2. “Did you know what the blue line (visual cue) coming from the vehicle was for?”, 3. “Did you notice the yellow frame highlight around the image?”) The result of each question is broken down into the audio activation method (T1 – T5 representing method 1 thru 5 as explained in Chapter 3). For example, 6 of 17 people asked who tested the zone method (T1) responded that they did know how to activate the audio description associated with each image in the exhibit. Subjects for the zone, visual cue and heading methods were asked question 1, users of all methods but the zone method were asked question 2 and users of the button press and the tractor beam methods were asked question 3.

Table 2. Overview of positive answers to the questions of the exit interview

	Question 1	Question 2	Question 3
T1	6/17	N/A	N/A
T2	6/14	12/14	N/A
T3	3/12	9/10	N/A
T4	N/A	10/13	7/14
T5	N/A	7/13	7/14
ALL	70%	76%	50%

Some reasons for not knowing what the blue line was for included, “I didn’t notice the blue line”, “I was too focused on driving” and some thought they saw the blue line prior to any sound (even though the blue line appears at the same time the audio begins to play). Only half of the people who gave answers to the interview noticed the yellow highlighted frame around the image that the vehicle was facing.

One person (an adult male) using the button press (T4) method mentioned having experience playing video games and had no problem noticing the yellow highlighted frame. Two people (both adult males) using the tractor beam (T5) method mentioned having no video game experience and a difficult time noticing the yellow highlighted frame. It was said to be too subtle for people who were not used to video game types of interaction with computers.

4.2.4 Informal Observations Made by Experimenter

Overall, 4th graders through teens appear to like, understand, and be able to easily use the trigger button. Teens and children also liked the video game aspect of shooting at things and having ultimate control. (Most do not seem to care about the audio in the zone, visual cue and heading methods, preferring more to drive through the environment fast). It appears the use of trigger buttons for these age groups encouraged them to slow down and even capture bits of information during their driving time.

Older people (Adult and Senior citizens) in general don’t seem to have as much interest in “shooting” things and have a hard enough time just learning how to drive the vehicle. It seems the extra button press for the older users just adds to their already taxed cognitive burden; however, the few that were successful at activating the audio with the button press paid attention to any audio they heard until it was over (avoiding driving or moving in many cases until the audio stopped).

There were various problems relating to the input device that were observed during the testing phase in this work: it was noted that younger children often complained of fatigue causing some to give up and leave the kiosk frustrated.

4.3 Study 2: Subjective Comparative Assessment

The subjective comparative assessment is used to find out the general level of preference each audio activation method had for users who tested all of the audio activation methods.

4.3.1 Subjects

Ten adult subjects (six female, four male) were solicited their help in evaluating the exhibit. Four of these were employees of the Seacoast Science Center, but not directly involved in the exhibit. The other six were visitors to the New England Aquarium. It was too difficult to involve children due to factors such as school trip schedules and impatient guardians.

4.3.2 Procedure

The subjects tested all five audio activation methods in varying order and ranked them, 1 being the best audio activation method and 5 being the worst. They gave comments during their experience and when their experience was completed, they answered the same questions as were in the objective testing exit interview. The average time for this testing phase was approximately 15 – 25 minutes per person.

4.3.3 Measures

The mean rankings ranged from 0 (least) to 5 (most) preferred audio activation method.

4.4 Results and Discussion of Study 2

The average rankings for each audio activation method used are illustrated in Figure 11.

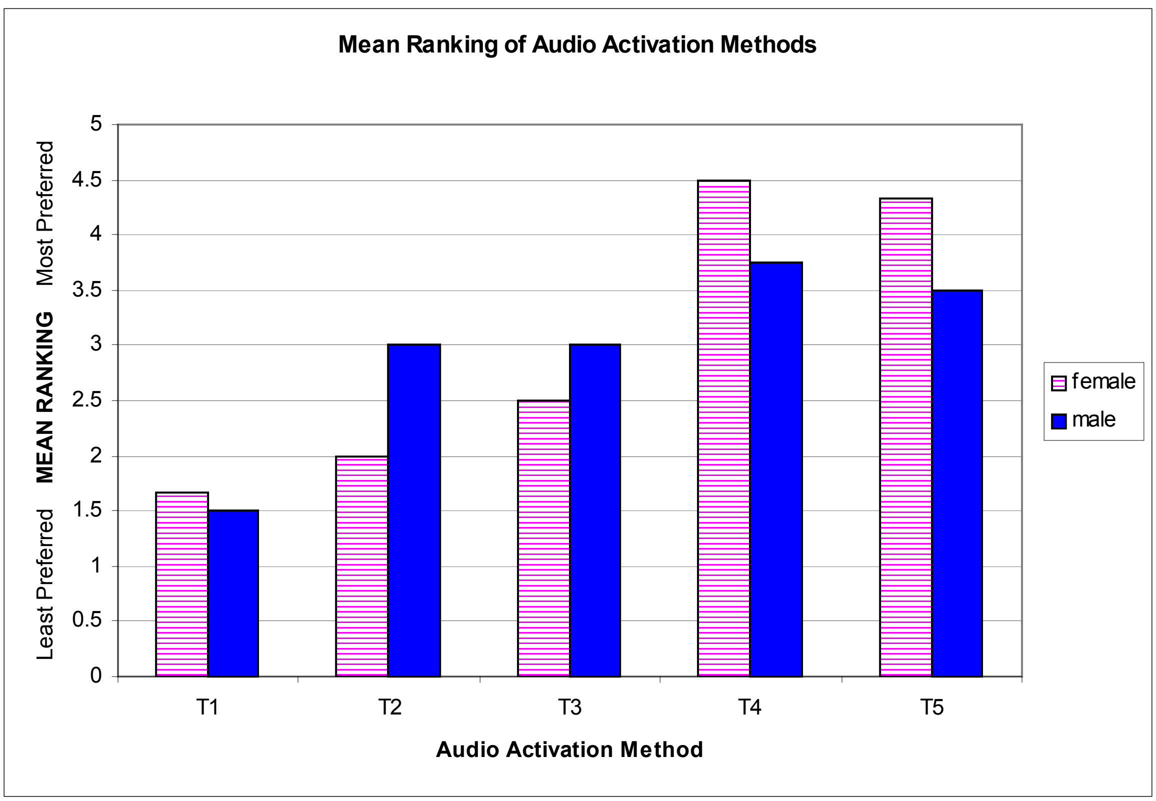

Figure 11. Mean rankings for each of the audio activation methods by gender.

The button press (T4) and tractor beam (T5) methods were rated overall best. Some of the reasons that were given include, "it gave the user more control", "more active participation", "you could choose your own picture when you wanted", and "I like to shoot the pictures". The tractor beam (T5) method was a close second and was chosen for similar reasons as the button press including comments such as, "it positioned you for good viewing" and "trigger use is good to be actively involved". However, from observation and recorded comments 4 of the 10 subjects didn't realize they had lost control of the vehicle during the tractor beam repositioning, therefore they didn't realize there was a difference between the button press (T4) and the tractor beam (T5) methods. But, when told of the difference between them they liked the idea of the tractor beam (T5) method better. Considering that both the button press (T4) and the tractor beam (T5) were the top ranking methods it supports the idea that active audio activation methods are more enjoyable for users. There were comments on how to improve the tractor beam if used in the general public including statements such as, "need to have an auditory explanation of how exactly to activate the audio", and "a constant headlight (visual cue) from the vehicle would be helpful".

The majority of the subjects agreed that the zone (T1) method was by far the worst of the five methods. They felt that the other four methods had much more to offer the user in terms of visual cue and active participation, even though the first one was easiest because there was less to do and see. Most preferred the blue line of the visual cue method (T2) to no visual cue for the zone method (T1). 4 of the 10 subjects felt there was no difference between activating audio for the visual cue (T2) and the heading (T3) methods; they failed to notice the use of heading for the third method.

Also, there was mention of the visual cues for signaling when the user was in line to activate audio with a button press (a yellow highlighted frame around the image) being too subtle to pick up on right away without explicit explanations.

4.5 General Discussion

The results of both the objective and subjective phases of testing indicate that the tractor beam (T5) method of audio clip activation is the best method overall for a virtual kiosk display that involves navigation and audio clip activation. This is considered the best method since the users rated it a very close second choice over all in the subjective assessment and since the objective results showed the majority of the users had higher times facing the images at audio activated locations.

Preferences for the button press (T4) and the tractor beam (T5) audio activation methods and results showing higher times facing the images for each of these methods implies that preference and times are correlated. This suggests that when the user has more control over accessing the content of the kiosk, the content that is eventually accessed has been accessed because there is genuine interest.

The subjective assessment study suggested that incorporating active methods of activating audio clips gives users a sense of control, which in turn allowed them to show specific interest in the environment they were apart of.

It is possible that the result for the average time facing images per activated audio clip according to gender (Fig. 10) is spurious due to the small number and age of female visitors using the tractor beam (T5) method (only 4 females total, 2 adults and 2 - 5th grade age). Eight females tested the button press (T4) method - 4 were in the adult category (3 teens and a senior citizen) and 4 were children (3 – 5th grade age and 1 was a 3rd grader.) Teens seemed to respond better to the use of trigger buttons, which seemed to encourage them to slow down and listen to the audio being played. This could be another explanation for why the button press (T4) method had better results than were expected and the tractor beam (T5) results were less than expected.

For the effective use of visual cues, the blue line was clearly a better visual cue (76% understood it) than the yellow highlighted frame (50% understood it). Practically everyone whose audio activation method included the use of the blue line noticed it and was aware of its purpose. The results were not the same with the yellow highlighted frame visual cue. Exit interview results suggest that the interface for the button press (T4) and the tractor beam (T5) were easier to use by video game players and they were better at picking up the frame visual cue (showing they had entered the activation zone).

Different methods provided different cues telling the user when they were in the audio activation zone. For the visual cue (T2) and heading methods (T3) a blue line (attached to the center of the image within the zone to the avatar vehicle) was used to indicate the audio activation zone (synchronized with playing the associated audio for the image). The button press (T4) and tractor beam (T5) methods used a highlighted yellow frame around the image to indicate the audio activation zone. As was noted, the highlighted yellow borders around the images were not noticed by many of the subjects using the button press (T4) and the tractor beam (T5) methods. Despite this, the button press (T4) and tractor beam (T5) methods had the longest times facing the images, according to the objective testing. These active methods were thus effectively penalized by having a poor visual cue, which strongly suggests that active methods are better.

These results suggest that visual cues can help to augment a given method (such as the difference between the zone (T1) and the visual cue (T2) methods) however, care needs to be takes as to their effectiveness when the cue is too subtle.

It is useful to note that “actions speak louder than words”, in other words some people said they liked a certain audio activation method better than another however their performance in the evaluation phase would often suggest contrary to what was verbalized.

CHAPTER 5

CONCLUSION

The ultimate goal of this thesis has been to achieve an enjoyable easy to use and informative exhibit. A more specific goal was to investigate how museum visitors should activate audio clips associated with images while actively maneuvering through a 3D computer-based world. To obtain a more objective measure this was translated into finding a method that would increase a visitor's time viewing images while corresponding audio commentary was played.

The audio activation methods were based, to different degrees, on the constructivist theory of learning, which states learning is more effective if the learner can actively construct links between the various aspects of the learning environment. Five methods were used to begin the activation of audio associated with the images: proximity of the avatar vehicle within a predefined zone, proximity in the zone using a visual cue (blue line), the proximity and heading of the vehicle within the zone using the visual cue, pressing a button within a zone while facing an image, and using a tractor beam to relocate the vehicle to a good viewing position which is also activated by pressing a button within a zone.

Some of the audio activation methods were less active without a clear cue (from the visitor's perspective) of when exactly audio could be played. For example, the first method, *proximity in the zone*, involved simple proximity, but visitors did not have a clear definition of the “zone” that triggered the audio activation for a given image. Since the zone was not visible to the visitor, the visitor did not have the ability to predict exactly when the audio would begin to play. However, the more active audio activation methods used a deliberate method of activating the audio by using a button press, making it more obvious when exactly audio could be played.

Due to these differences in the level of activity needed to activate audio, there are trade-offs between the less active and highly interactive forms of audio activation methods. The first trade off is ease-of-use versus confusion of the audio activation zones; the visitor can easily use the less active audio activation methods yet s/he can’t pinpoint the exact moment of activation and perhaps may be confused as a result. Another trade-off is higher cognitive load versus more control; the more active audio activation methods require more direct actions and may demand more cognitive resources, but give the visitor more control over when audio activations occur. These methods may also have been harder to learn.

Although it was not initially clear which would be the best method, the evaluation that was done clearly showed that active methods of audio activation were more effective than passive methods. The key findings indicated that although the active audio activation methods of the button press and the tractor beam were more complicated for the user, they were generally preferred and led to visitors spending more time in front of audio activated images. This preference of active methods over the more passive methods of audio activation supports constructivist theories. The tractor beam method in particular gave the user full control over when they wanted the audio to activate, yet helped less skilled users to position the vehicle in a better location for viewing the image.

These results led to a decision to use the tractor beam audio activation method in the kiosk at the Seacoast Science Center. However, the subjective testing of the visitors suggested a few alterations to make the tractor beam method even more effective: better visual and auditory cues for initially calling attention to the user as to when they are able to "shoot" the image, and a better auditory explanation of the behavior of the tractor beam. A better explanation should include adding an audio commentary the first time the button is used to activate an audio clip such as, "You are now being repositioned to a better viewing location." These changes have been implemented for the version used in the Seacoast Science Center.

The updated version of GeoZui3D Explorer 1 also includes an interactive tutorial to help the user become familiar with the input device. The need for this became evident during subjective testing. The new instructions are far more interactive; including, for example, visitors being asked to "move forward" by pushing the input device away from them. The system waits for the action to take place and gives a positive response of "good" then moves on to the next instruction. If no movement or the wrong movement occurs then the instruction is repeated.

The Seacoast Science Center is undergoing renovations that are scheduled for completion near the end of April 2004. At the time of completion the general public will then be invited back for regular operating hours. The improved version of the GeoZui3D Explorer 1 exhibit will be beta-tested with the public for the first month of operation in order to discover and resolve usability issues.

5.1 Future Work

As noted in section 4.2.3 – Informal Observations Made by Experimenter, there were problems with the yoke. Also the manufacturer of the yoke, possibly due to its shortcomings, has discontinued the input device chosen for the Seacoast Science Center. It might be beneficial to use another two-dimensional input device (a joystick for example) to control turning the avatar side to side and tilting it up and down with another device (i.e. a throttle type lever) to handle the third dimension of forward motion.

This study focused on the *behavior* of visitors within a highly interactive 3D kiosk. A more in-depth study of the *retention* of the content within an exhibit would be a possible direction for future work. Such a study could include the influence of active and passive learning within these types of environments.

5.2 Summary

Edutainment is defined as "the act of learning through a medium that both educates and entertains" [Mifflin, 2000]. 3D interactive kiosk environments hold promise for edutainment. This fun method of introducing new topics for learning could encourage students to actively participate in the learning process, thereby furthering their understanding of the material covered and helping to increase their level of interest.

Because the objective of a public kiosk in a museum or science center is to entertain as well as educate, a suitable method of activating audio while navigating would integrate elements of both. Allowing the interface to be highly interactive adds to the entertaining value of the kiosk, and using a method that encourages users to face images longer while listening to the accompanying audio supports the educational goal of a kiosk. This constructivist approach in virtual reality kiosk environments helps to augment the user's overall experience.

The novel tractor beam method of audio clip activation proved to be the best. The tractor beam is activated when the user is facing an image within a predetermined zone (with entry in the zone signaled by a yellow highlighted frame around the image) and s/he presses the trigger button. At this point a ray (blue line) links the avatar with the center of the image and s/he temporarily loses control of the avatar while it is smoothly repositioned to a central position in front of the image. When the avatar is at the appropriate location, the audio clip begins to play and control is returned back to the user. The properties of the tractor beam inspire users to linger in front of the images longer than the other audio activation methods tested in this study, thus making it the method of choice for the exhibit at the Seacoast Science Center.

REFERENCES

Baddeley, A. D. (1986). *Working Memory*. Oxford, Oxford University Press.

Barbieri, T., F. Garzotto, et al. (2001). From dust to stardust: a Collaborative 3D Virtual Museum of Computer Science in *Proceedings ICHIM 01*, Milano, Italy.

Bricken, M. and C. M. Byrnes (1993). Summer students in virtual reality: a pilot study on educational applications of virtual reality technology. *Virtual reality: Applications and explorations*. A. Wexelblat. Boston, MA, Academic**:** 199-218.

Brooks JR., F. P. (1988). Grasping Reality Through Illusion: Interactive Graphics Serving Science. *Proceedings of the Fifth Conference on Computers and Human Interaction*, ACM.

Brown, J. S., A. Collins, et al. (1996). Situated cognition and culture of learning. *Situated learning perspectives*. H. McLellan. New Jersey, Educational Technology Publications.

Clark, J. M. and A. Paivio (1991). Dual Coding Theory and Education. *Educational Psychology Review* **3**(3): 149-210.

Cohen, R. and D. L. Weatherford (1981). The Effects of Barriers on Spatial Representation. *Child Development* **52**: 1087-1090.

Corbetta, M. and G. L. Shulman (2002). Control of Goal-Directed and Stimulus-Driven Attention in the Brain. *Nature Reviews Neuroscience*. **3:** 201-215.

Dick, W. and L. Cary (1990). *The Systematic Design of Instruction*, Harper Collins.

Duffy, T. M. and D. J. Cunningham (1996). Constructivism: Implications for design and delivery of instruction. *Handbook of research for educational communications and technology.* D. Jonassen. New York, Macmillan.

Feldman, A. and L. Acredolo (1979). The Effect of Active versus Passive Exploration on Memory for Spatial Location in Children. *Child Development* **50**: 698-704.

Fishbein, H. D., T. Echart, et al. (1990). Learners' Questions and Comprehension in a Tutoring Setting. *Journal of Educational Psychology* **82**(1): 163-170.

Gaitatzes, A., D. Christopoulos, et al. (2001). *Virtual Reality Interfaces for the Broad Public.*

Guide-Man (2002). Audio Guides, Ophrys Systemes**:** 1-10. http://www.ophrys.net

Hanke, M. A. (1992). Explore the Fort at Mashantucket, Design Division, Inc., of New York. **2003**. http://www.pequotmuseum.org

Hazen, N. L. (1982). Spatial Exploration and Spatial Knowledge: Individual and Developmental Differences in Very Young Children. *Child Development* **53**: 826-833.

Hibbard, W. and D. Santek (1989). Interactivity is the key. *Proceedings of the Chapel Hill Workshop on Volume Visualization*, Chapel Hill, North Carolina, U.S., ACM Press, New York, NY, USA.

Hinckley, K., R. Paush, et al. (1994). A survey of design issues in spatial input. *7th annual ACM symposium on User interface software and technology*, Marina del Rey, California, USA.

Hitch, G. J. (1987). Working memory. *Applying Cognitive Psychology To User-Interface Design*. M. M. Gardiner and B. Christie. New York, John Wiley & Sons**:** 120-121.

Jonassen, D. and L. Rohrer-Murphy (1999). Activity Theory as a Framework for Designing Constructivist Learning Environments. *Educational Technology Reaserach and Development* **47**(1): 61-79.

Johnson, W. L., J. W. Rickel, et al. (2000). Animated Pedagogical Agents: Face-to-Face Interaction in Interactive learning Environments. *Interanational Journal of Artificial Intelligence in Education* **11**: 47-78.

Kearsley, G. and B. Shneiderman (1998). Engagement theory: A framework for technology-based teaching and learning. *Educational Technology* **38**(5): 20-23.

Kranstedt, A., P. Kuhnlein, et al. (2003). Deixis in Multimodal Human Computer Interaction: An Interdisciplinary Approach. University of Bielefeld, Germany, *Gesture Workshop*, Genova, Italy, Springer-Verlag.

Larson, B. (1996). Boats Afloat Exhibit for the Children's Museum, Boston. http://www.bradlarson.com/projects1.htm

Lave, J. and E. Wenger (1991). *Situated Learning: legitimate peripheral participation.* Cambridge, MA, Cambridge University Press.

Mifflin, H. (2000). The American Heritage Dictionary of the English Language, Houghton Mifflin Company. **2004**.

Kelly, R. (1993). VR and the educational frontier. *Virtual Reality Special Report*, fall.

Mayer, R. E. and R. B. Anderson (1992). The Instructive Animation: Helping Students Build Connections Between Words and Pictures in Multimedia Learning. *Journal of Educational Psychology* **84**(4): 444-452.

Mayer, R. E. and V. K. Sims (1994). For whom is a picture worth a thousand words? Extensions of a dual-coding theory of multimedia learning. *Journal of Educational Psychology* **86**(3): 389-401.

Mayer, R. E. and R. Moreno (1998). A split-attention effect in multimedia learning: Evidence for dual processing systems in working memory. *Educational Psychology* **90**: 312-320.

Mayer, R. E. and R. Moreno (1998a). *A Cognitive Theory of Multimedia Learning: Implications for Design Principles*. ACM SIGCHI conference on Human Factors in Computing Systems, Los Angeles, CA.

Mayer, R. E., R. Moreno, et al. (1999). Maximizing constructivist learning from multimedia communications by minimizing cognitive load. *Journal of Educational Psychology* **91**(4): 638-643.

Melanson, B., J. Kelso, et al. (2001). Effects of Active Exploration and Passive Observation on Spatial Learning in a CAVE, Department of Computer Science, Virginia Tech**:** 1-11.

MacDonald, L. and J. Vince, Eds. (1994). *Interacting with Virtual Environments*, John Wiley & Sons Ltd.

Oviatt, S., A. DeAngeli, et al. (1997). Integration and Synchronization of Input Modes during Multimodal Human-Computer Interaction. *CHI 97*, Atlanta, GA, ACM.

Paivio, A. (1986). *Mental representations: A dual coding approach*. New York, Oxford University Press.

Ressler, S. and Q. Wang (1998). Making VRML accessible for people with disabilities. *ACM SIGCAPH Conference on Assistive Technologies, Proceedings of the third international ACM conference on Assistive technologies*, Marina del Rey, CA, ACM Press New York, NY.

Schaller, D. T. and S. Allison-Bunnell (2003). Practicing What We Teach: how learning theory can guide development of online educational activities. *The Museums and the Web 2003 conference*, Archives and Museum Informatics.

Spaargaren, O., J. E. Wien, et al. (2002). An Interactive Virtual Soil Museum: A Pilot Project of the International Soil Reference and Information Centre (ISRIC) with Focus on Arid Soils. *International Arid Lands Consortium (IALC)*, Tucson, Arizona, Arid Lands Newsletter.

Stock, O. and M. Zancanaro (2002). *Intelligent Interactive Information Presentation for Cultural Tourism*. Invited talk at the International Workshop on Natural, Intelligent and Effective Interaction in Multimodal Dialogue Systems, Copenhagen, Denmark.

Tellevik, J. M., B. Elmerskog, et al. (2003). Observing Directed Attention in Mobility. Oslo, Norway, Tambartun National Resource Center of special education of the visually impaired. Department of Special Education, University of Oslo, Norway**:** 6.

Ware, C. (2000). *Information Visualization: Perception for Design*. San Francisco, Morgan Kaufmann Publishers.

Ware, C., M. Plumlee, R. Arsenault, L. A. Mayer, S. Smith, and D. House (2001). GeoZui3D: Data Fusion for Interpreting Oceanographic Data, *Proceedings Oceans 2001*.

Wilson, P. N., N. Foreman, et al. (1997). Active Versus Passive Processing of Spatial Information in a Computer-Simulated Environment. *Ecological Psychology* **9**(3): 207-222.

Wilson, P. N. (1999). Active exploration of a virtual environment does not promote orientation or memory for objects. *Environment and Behavior* **31**(6): 752-763.

APPENDIX A

DATA FROM LENGTH OF STAY AT EXHIBITS ASSESSMENT

Visitors of a museum have a short amount of time, some are from out of town and may not visit the museum ever again, and this may mean they will rush through the exhibits in order to take in everything they can. Sometimes, it is because a school group has come through and the teacher in charge is only allowing a short amount of time per exhibit. When building an interactive computer-based interface for a museum exhibit, the design must take into consideration the estimated amount of time a visitor is expected to spend at the exhibit. Having an expected time per kiosk user, will allow for the design to have reasonable guidelines for how long the interface should take to be learned.

The goal of this assessment is to find out through observed visits to the Boston Museum of Science the average time a typical patron of the museum spends on specific exhibits. The objective is to find an exhibit that has computer-based interactive elements. The categories that were assessed include exhibits that are more involved with a specific objective involved or task to complete, exhibits without a specific objective that are low in interaction but high in observation, exhibits that are highly interactive with a creative influence and typical interactive educational computer games.

Exhibit 1: Virtual Maze (objective involved)

Bill Keays and Ron MacNeil created the virtual maze (Figure 9), located in the current science and technology center of the Boston Museum of Science. The idea behind this was to explore new interfaces between people, computers and their environment. It is based on a familiar mechanical board game that involves tilting the board in various directions in order to guide a small steel ball through the maze. Throughout the maze there are holes that the ball could fall into, thus making the game even more challenging. The virtual maze, set up in the same fashion as the mechanical board game, is the projection of a computer screen on the floor of the museum spanning a distance of approximately 5' by 6'. A camera above the maze tracks the users movements and calculates the tilt in the board on the floor, based on the users body mass and their location. Thus giving the illusion that the user is actually tilting the board each time they step on a different spot on the floor.

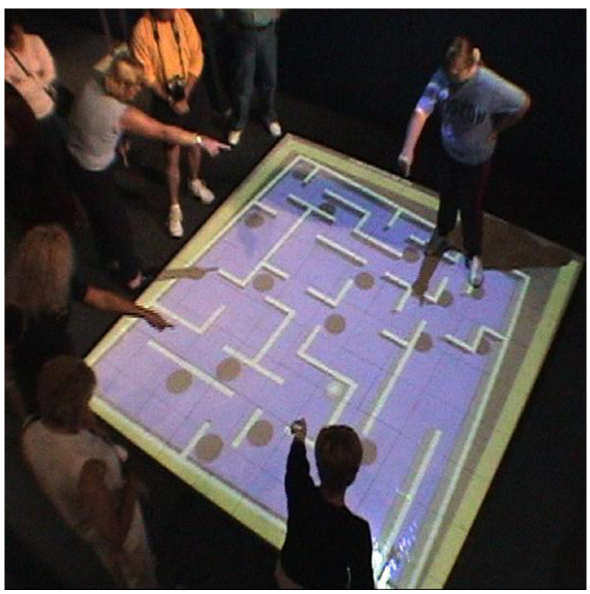

Figure 1. Virtual Maze, Boston Museum of Science

There were no initial instructions on how to work the game at the exhibit and only a few went near the associated computer display in a plexi-glass box that was set up nearby. This was a very popular exhibit with many onlookers and children waiting in line to play. Some children didn't participate do to the "stage fright" that was associated with so many people watching. Some wanted to play so badly that they walked away and came back later when less people were around. Many children who did play the game were forced to leave by their accompanying adult after a minute or so and many were joined by at least one other child while playing, sometimes forcing one or more children off the game prior to their completing the task. Not many made it to the end of the maze but most tried at least two times. The major age group that visited this exhibit was 12 and under and only 1 teen out of 20 people tried it out. Only one adult attempted the virtual maze but left shortly after stepping on it, as if he shouldn't be occupying it. Nine of the twenty visitors timed were female, showing this had equal appeal to males. 16 people looked at the exhibit and left without trying it. From observation, it was noted that the average length of time for each participant at this exhibit was approximately 1 minute 14 seconds, and all seemed to learn it quickly, within 1 second of trying it.

Exhibit 2: Minimally Invasive Surgery (objective involved)

The Current Science and Technology Center in the Boston Museum of Science is showcasing a display on the use of minimally invasive surgery. This display is set up similar to the picture below (Figure 10), however it has been somewhat simplified. There is a covered box with two robotic type arms sticking out of it. The arms are devices used, one in each hand, to thread a small rope through three metal screws with eyeholes at the end. Each arm control was placed approximately the shoulder width of an adult apart. There is a photo of an intestine beneath the screws that gives the impression that threading the rope through the eyes of the screws is following the path of the intestine. Inside the box is a camera that projects the image of the inside of the box to a computer screen nearby. There is a peek-a-boo flap in-between the two robotic arm handles that allow the user to peek into the box, if it is too difficult to watch the work they are doing on the computer screen to the right of them. The controls involved gripping handles with each hand and squeezing them to pinch the end of the grippers inside the box and releasing them opened up the grip.

Figure 2. Minimally Invasive Surgery Exhibit, Boston Museum of Science.

There was a 2-3-paragraph placard near the controls of the exhibit containing an explanation of the exhibit and the objective that most adults seemed to read for at least 5 seconds. Children didn't appear to even look at the directions. Most of the adults that approached this exhibit seemed interested and attempted whole-heartedly to accomplish the task. The children on the other hand, played with the controls and seemed to give up on the task at hand after a few tries. No one observed finished stringing the rope through all the eyehole. Some of the children shook their hands off after attempting the controls for a little while as if to indicate their hands were fatigued. The major age group that visited this exhibit was 12 and under. There was only 1 teen, 1 senior and 6 adults out of 20 people who tried it out. One of the twenty visitors timed was female, showing this had a much higher appeal to males. 13 people looked at the exhibit and left without trying it. From observation, it was noted that the average length of time for each participant at this exhibit was approximately 1 minute 32 seconds, and all seemed to learn it in less than 10 seconds.

<u>Exhibit 3: Bioscanner (no objective involved)</u>

In the Seeing the Unseen section of the Green Wing in the Boston Museum of Science the Bioscanner is on display. This display is used to take a look at life “up close and personal”. With the aid of three different controls on the control panel, the user has the opportunity to control the zoom, focus and direction of a camera that peers into the world of creatures and textures. The zoom and the focus controls are the similar to one another, both require pushing the lever forward or backward for a zoom in or out, or more or less focus, respectively. The directional control uses a joystick input device that allows for freedom to look around in all directions on a two-dimensional plane. There is no goal to accomplish for this exhibit, just exploration of the world inside a glass fish tank.

There was a placard near the controls of the exhibit containing a short introduction of the exhibit and three points for the operating instructions (one for each of the controls). No one noticeably read the directions. Most of the children that approached this exhibit shared their findings with someone in their party. They also seem to be bored and move on after figuring out the controls and inspecting a few items. The major age group that visited this exhibit was 12 and under. There were six teens and one adult out of 20 people who tried it out. 5 of the 20 visitors timed were female, showing this had a much higher appeal to males. All who passed this exhibit seemed to try it out. From observation, it was noted that the average length of time for each participant at this exhibit was approximately 32 seconds, and all seemed to learn it in less than 1-2 seconds.

Exhibit 4: Disappearing Act (no objective involved)

In the Seeing the Unseen section of the Green Wing in the Boston Museum of Science there is a display titled "Disappearing Act". This display is used to show how animals can blend into their environment by not moving, the second they begin to move their cover is gone. This computer based exhibit contains 3-4 different optical illusion type screens. With the aid of two buttons on the control panel, the user can stop the movement of the fish to watch it disappear into its surroundings or change the scene to a new fish and background. There is no real goal to accomplish for this exhibit, just notice the fish appear and disappear from its environment. Figure 11 is an example of what the initial screen looks like when the stop button is pressed.

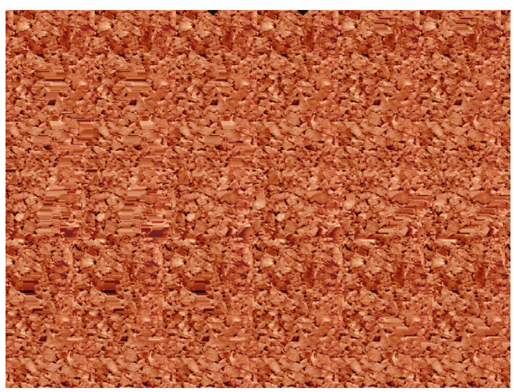

Figure 3. Disappearing Act screen of exhibit, Boston Museum of Science

There was a placard near the controls of the exhibit containing a long explanation of the illusion of the exhibit. First was an explanation of the control buttons (two sentences), then 6-8 sentences on what is happening with the illusion, then about three more sentences on "so what?" regarding why this is important. Very few noticeably read the information. Most of the people that approached this exhibit seem to be bored and move on after figuring out the controls and inspecting the first screen. Many of the children at this exhibit just pushed the buttons while looking elsewhere. All the age groups evenly visited this exhibit. There were 7 children, 7 teens and 5 adults and 1 senior out of 20 people who tried it out. Nine of the twenty visitors timed were female, showing this had an equal appeal to males. Only one adult looked at it and walked away without trying it out. From observation, it was noted that the average length of time for each participant at this exhibit was approximately 35 seconds, and no one took any extra time to learn it.

Exhibit 5: Virtual Fish Tank (creative)

The Virtual Fish Tank, built by Nearlife, Inc. in conjunction with the Museum of Science, is the first Computer Museum exhibit to open at the Science Park. The Virtual Fish Tank was built with the goal to "stimulate the interest in and further understanding of science and technology and their importance for individuals and for society." It is designed as an interactive computer-based exhibit that allows the participants to create their own fish both physically and behaviorally. The choices involved in the fish creation are divided into three categories: Food Chain, Interests, and Depth. In the Food Chain category are choices for eyes (fear) ranging from terrified to calm and the mouth (hunger) ranging from full and starved. In the Interests category are choices for the shine (bright objects) ranging from like to dislike, fins (bubbles) ranging from annoy to enjoy, and smile (mouse clicks) ranging from shy to friendly. In the Depth category the choice is the body shape (depth range) ranging from the top of the tank to the bottom of the tank. Figure 12 depicts the completed look of three fish using this design process.

Figure 4. Virtual Fish created at the exhibit

The exhibit is set up with 4-5 touch computers screens set up for the fish design center, facing a very large wall projection of the combined 24,000 gallon tank. Each screen had an associated physical "tube" on the side of it connected to the top of the tank to emulate the path the fish would follow to get to the combined virtual fish tank. As the fish entered the virtual fish tank, from the individual design center screen, through the release tube, the participant could watch it drop out of the open end of the tube. Afterwards the participant could follow their fish, observe its behavior and even release "food" for all the fish. A large description of the exhibit was placed at the entrance of the entire exhibit, and each individual design center had a short description of the design process. Most people seemed to understand how to design and release the fish into the tank after reading the posted instructions. Everyone appeared excited with the exhibit and engaged in the creation of their fish. Excitement was evident as the observed fish acted on the pre-programmed behaviors and especially when another fish ate a fish.

Teenagers were much more attracted to this exhibit than any of the exhibits tested in these evaluations. Overall, there were 2 adults, 7 teens, and 11 children visited and participated in this exhibit. Only a few seniors and adult looked at the exhibit without trying to create their own fish. 9 of the visitors were female, showing an equal appeal for this exhibit from each gender. This exhibit seemed to be quickly understood and participants were fully functioning within 5-30 seconds. From observation, the average time was 5 minutes 31 seconds, which seemed to be influenced mostly by the creative element involved in this exhibit.

Exhibit 6: Motion Illusion (creative)

Using the popular Powerpuff characters (see Figure 13) from the Cartoon network, the participants of this exhibit can make their own movie. There is one poster board set up to explain the math behind the exhibit that states: "Speed is distance/time. Record a series of stationary scenes (frames). You create the illusion of speed by moving the props between frames. Press play and your movie runs at 20 frames per second (FPS). 2 inches/10 frames X 20 frames/1 sec = 4 inches/second." On the other side of the display for the exhibit there is a detail list of items to do in order to successfully create a Powerpuff movie. The background for the props was a spin-top table that had a cartoon type landscape on it, which added to the illusion of the props moving through the scene. The camera was positioned in an overhead view as to capture the scene directly below it, and a monitor was placed directly in front of the work area showing both real time action and the play back of the recorded series of events.

Figure 5. Powerpuff girls used as props for the Motion Illusion exhibit. Buttercup, Blossom and Bubbles from left to right.

The basic controls for this exhibit were three buttons; one for recording the current frame, one for deleting all the frames and starting over, and one to playback the series that had been recorded. Most people that viewed this exhibit seemed fascinated with it. They all tried taking at least one picture of the scene they made, most played it back as well. Some got really creative making the movie, then called friends over to see the results. On occasion some walked away then returned to finish up or to try something new. Children were the primary visitors of this exhibit with only 2 teens and 2 adults visiting out of 20 people observed. Only 5 females visited this exhibit in the time observed which was outside of the expectation considering the props seemed more oriented towards younger females. There was an obvious learning time of a minimum of 5 seconds to the entirety of the time spend on the exhibit. On average people who left the exhibit before one minute didn't seem to really understand how to utilize the exhibit, of these most appeared to not read the directions that were posted. For those that did eventually make a "movie" learned either by observing the previous visitor or within 30 seconds. From observation the average time spent at this exhibit was 2 minutes 14 seconds.

<u>Exhibit 7: Stop That Train (computer game)</u>

One of Japan's all time most popular video games called Let's Go by train (Figure 14 – Japanese name for the game is Densya de Go!). This exhibit was set up for an acceleration/deceleration hands on experience. The game is set up so the participant takes on the role of the train conductor. Part of the role involves picking up passengers at different stops and getting them to their destination on time. There are four different train lines to choose from each with a different environment, challenging conditions and control responses. Many factors are involved in playing a successful game as the conductor; you can't go too fast or you will catch up to the train in front of you, you can't go too slow or you will be late and be deducted points for every second you are late, you must prepare for things crossing the track, track changes, appropriate signaling, etc.

Figure 6. Your perspective in the game is looking out the train's windshield.

The areas on the screen seen are labeled by numbers that mean the following; 1) Shows throttle position. 2) Time due at station. 3) "Present" time. 4) Points remaining. 5) To the left, the train type, below, a picture of the train. 6) Names of stations operating between 7) Position of brake handle, 8) This symbol means that there is no posted speed restriction present. 9) Distance to station stop marker, which can be shut off to make game harder! 10) Speedometer in km/h.

The controls for this game consist of separate acceleration and deceleration joysticks that are pushed forward and backward, a large start button in between the joysticks, a large button for changing the viewpoint, a large button for signaling the horn, and a camera joystick. Not many people seemed to read the instruction posted on the side of the computer screen to find out exactly how to work the controls. The only video instructions that were in English were on the startup screen of the video game, otherwise all of the writing on the screen was in Japanese.

The Let's Go by train computer game attracted 5 adults and the rest 20 people observed were children. Only 2 of the participants were female, which showed that males had a much greater interest in this exhibit. From observation, the average time was shown to be 2 minute 39 seconds.

<u>Exhibit 8: Children's Video Games (computer game)</u>

Two children's educational video games were tested for this observational time trial. The first one a child's game for the ages group from 4-7 called Freddi Fish, and the other one, the Oregon Trail, targeted for children ages 9 and up. Ten people total were observed for the following two games. Because each of these games is PC based the input device for both of them were the mouse and keyboard.

Freddi Fish is a game that involves collecting clues in order to be able to "help Freddi Fish catch the thieves who stole cousin Calico's herd of hogfish". The game includes bright bold cartoon characters (see Figure 15) that come to life while explaining the story at hand. Some of the educational aspects involved in this game are; critical thinking, listening and memory skills, and a reinforcement of good social skills. Exploration is involved to find clues and follow the trail of the thieves.

Figure 7. Freddi Fish from the children's software exhibit.

The Freddi Fish game attracted four children and one teen of the four people observed two were female, showing an equal interest for both genders. The age of the children playing this game made it a bit tricky to determine how well they understood the game and if they were playing correctly, but they seemed to be able to navigate from screen to screen within a few seconds of starting the game. From observation, the average time was shown to be 7 minutes 53 seconds.

The Oregon Trail is a game that involves walking in the steps of pioneers and learning from the same adventures they had. This game also has cartoon like characters (see Figure 16) and 3-D graphics. Some of the educational benefits of this game include real-life decision-making and problem-solving skills. The player chooses his wagon party, the supplies needed, they read maps, plan the route and guide the team through the wilderness. They must over-come sickness and starvation, traverse raging rivers, and avoid buffalo stampedes.

Figure 8. Characters from the Oregon Trail in the children's software exhibit.

The Oregon Trail game attracted 5 older children (pre-teens) and one teenager of the six people observed. Again exactly half were female, showing an equal interest for both genders. The learning time appeared to range from 5-10 seconds, however, most appeared to be comfortable with the program as if they had previous experience with it. From observation, the average time was shown to be 3 minutes 5 seconds.

The overall time spent for the combined PC games in this section were 9 minutes 18 seconds for the ten people observed.

Results

Table 3 shows the summary of the statistics gathered:

Table 1. Overview of the statistics collected from observational visit to the Boston Museum of Science.

Exhibit	Category	Female	Senior	Adult	Teen	Children	Time to Learn (seconds)	Average Time Spent
Virtual Maze	Objective	9	0	0	1	19	0-1	1 min 14 s
Surgery	Objective	1	1	6	1	12	5-10	1 min 32 s
BioScan	Observation	5	0	1	6	13	0-2	32 s
Disappear	Observation	9	1	5	7	7	0-1	35 s
Fish Tank	Creative	9	0	2	7	11	5-30	5 min 31 s
Make Movie	Creative	5	0	2	2	16	5-50	2 min 14 s
Stop Train	Comp Game	2	0	5	0	15	30-60+	2 min 39 s
PC Games	Comp Game	5	0	0	2	8	0-10	9 min 18 s

The observation class of exhibit (with minimal interaction) yielded the shortest visit times overall. The longest times reported were the PC Game exhibits with average times of over 9 minutes, and the highly creative exhibits that had novel ideas and interfaces (with average times of 5 and ½ minutes).

The Virtual Fish Tank showed that a highly interactive, goal-based, computer-based exhibit can capture the attention of a variety of people and keep them there over 5 minutes. It also implied that an efficiently designed interface that is both simple, intuitive, and briefly (yet thoroughly) explained can be learned in well under a minute, allowing for full enjoyment of the content of the exhibit.

APPENDIX B

IRB HUMAN SUBJECTS IN RESEARCH APPROVAL

UNIVERSITY OF NEW HAMPSHIRE

Office of Sponsored Research
Service Building
51 College Road
Durham, New Hampshire
03824-3585
(603) 862-3564 FAX

LAST NAME	*Sullivan*	**FIRST NAME**	*Briana*
DEPT	*Computer Science Dept/Center for Coastal & Ocean Mapping, Chase Ocean Engineering Lab*	**APP'L DATE**	*6/4/2003*
OFF-CAMPUS ADDRESS (if applicable)	*Computer Science/Center for Coastal & Ocean Mapping Chase Ocean Engineering Lab*	**IRB #**	*2971*
		REVIEW LEVEL	*EXE*
		DATE OF NOTICE	*6/4/2003*
PROJECT TITLE	*Flying Interfaces for Museum Exhibits*		

The Institutional Review Board (IRB) for the Protection of Human Subjects in Research has reviewed and approved the protocol for your study as Exempt as described in Federal Regulations 45 CFR 46, Subsection 101 (b), category 2.
The IRB made the following comment(s) or recommendation(s). They are not contingencies, and do not require a formal response from you unless otherwise noted.

- Approval at this time is only for observation at Boston Museum of Science (research protocol #1). IRB review of interviews (research protocol #2) will take place next week at which time the investigator will be informed of the outcome.

Approval is granted to conduct the study as described in your protocol. **Prior to implementing any changes in your protocol, you must submit them to the IRB for review and receive written, unconditional approval. If you experience any unusual or unanticipated results with regard to the participation of human subjects, report such events to this office within one working day of occurrence.** Upon completion of your study, please complete the enclosed pink Exempt Study Final Report form and return it to this office along with a report of your findings.

The protection of human subjects in your study is an ongoing process for which you hold primary responsibility. In receiving IRB approval for your protocol, you agree to conduct the study in accordance with the ethical principles and guidelines for the protection of human subjects in research, as described in the following three reports: Belmont Report; Title 45, Code of Federal Regulations, Part 46; and UNH's Federalwide Assurance of Protection of Human Subjects. The full text of these documents is available on the Office of Sponsored Research (OSR) website at http://www.unh.edu/osr/compliance/Regulatory_Compliance.html and by request from OSR.

If you have questions or concerns about your study or this approval, please feel free to contact me at 862-2003. Please refer to the IRB # above in all correspondence related to this study. The IRB wishes you success with your research.

For the IRB,

Julie F. Simpson

Julie F. Simpson
Regulatory Compliance Manager

cc: File
Colin Ware, Advisor
Matthew Plumlee, Computer Science

UNIVERSITY OF NEW HAMPSHIRE

Office of Sponsored Research
Service Building
51 College Road
Durham, New Hampshire 03824-3585
(603) 862-3564 FAX

LAST NAME	*Sullivan*	**FIRST NAME**	*Briana*
DEPT	*Computer Science Dept/Center for Coastal & Ocean*	**APP'L DATE**	*6/11/2003*
OFF-CAMPUS ADDRESS (if applicable)	*Computer Science/Center for Coastal & Ocean Mapping Chase Ocean Engineering Lab*	**IRB #**	*2971*
		REVIEW LEVEL	*EXP*
		DATE OF NOTICE	*6/13/2003*
PROJECT TITLE	*Flying Interfaces for Museum Exhibits*		

The Institutional Review Board (IRB) for the Protection of Human Subjects in Research reviewed and approved the protocol for your study as Expedited as described in Federal Regulations 45 CFR 46, Subsection 110 (b) (1)category 7 . The IRB made the following comment(s) or recommendation(s). They are not contingencies, and do not require a formal response from you unless otherwise noted.

- This approval is for interview protocol.

Approval for this protocol expires one year from the approval date above. At the end of the approval period you will be asked to submit a report for this study with regard to the involvement of human subjects. If your study is still active, you may apply for extension of IRB approval through this office.

The protection of human subjects in your study is an ongoing process for which you hold primary responsibility. In receiving IRB approval for your protocol, you agree to conduct the study in accordance with the ethical principles and guidelines for the protection of human subjects in research, as described in the following three reports: Belmont Report; Title 45, Code of Federal Regulations, Part 46; and UNH's Federalwide Assurance of Protection of Human Subjects. The full text of these documents is available on the Office of Sponsored Research (OSR) website at http://www.unh.edu/osr/compliance/Regulatory_Compliance.html and by request from OSR.

Changes in your protocol must be submitted to the IRB for review and approval prior to their implementation. If you experience any unusual or unanticipated results with regard to the participation of human subjects, report such events to this office within one working day of occurrence. If you have questions or concerns about your study or this approval, please feel free to contact me at 862-2003. Please refer to the IRB # above in all correspondence related to this study. The IRB wishes you success with your research.

For the IRB,

Julie F. Simpson

Julie F. Simpson
Regulatory Compliance Manager

cc: File
Colin Ware, Advisor
Matthew Plumlee, Computer Science