Linking Audio And Visual Information While Navigating In A Virtual Reality Kiosk Display

***Abstract*—** 3D interactive virtual reality museum exhibits should be easy to use, entertaining and informative. If the interface is intuitive, it will allow the user more time to learn the educational content of the exhibit. This paper is concerned with interface issues concerning activating audio descriptions of images in such exhibits while the user is navigating. Five methods for activating audio descriptions were implemented and evaluated to find the most effective. These range roughly on a passive-active continuum; with the more passive methods an audio explanation was triggered by simple proximity to an image of interest and the more active methods involved users orienting themselves and pressing a button to start the audio. In the most elaborate method, once the visitor had pressed a trigger button, the system initiated a "tractor-beam" that animated the viewpoint to a location in front of and facing the image of interest before starting the audio. The results of this research suggest that the more active methods were both preferred and more effective in getting visitors to face objects of interest while audio played. The tractor-beam method was best overall and implemented in a museum exhibit.

***Index Terms*— multimedia, virtual reality, educational software, kiosk**

I. INTRODUCTION

Modern computer technology has made possible 3D interactive public kiosks that provide the user with a multi-media rich environment that may include text, graphics, images, sound-clips, video, and animations. Often these environments allow the user to interactively select content and navigate through the 3D space to retrieve information;

however, the navigation task may distract the user from this information. Ideally, the user should enjoy the benefits of these kiosks without sacrificing the ability to acquire the information they contain. Developing these types of interactive environments is a complex task due to the specific requirements of kiosks. That is, they should be exceptionally easy to use, as they must be proficiently operated within a few minutes; they should be self-explanatory as there are no human helpers to interact with; and they should engage users with interesting content so their experience will be a memorable one.

This paper 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-Bunnell, 2003; Travis, Watson, Atyeo, 1994). This requires finding an appropriate method for activating audio descriptions that is simple to learn and use.

This research also had a practical goal. By obtaining a contract from the New Hampshire Seacoast Science Center, it was possible to design and build the interface for a 3D kiosk; with the intent to inform the public about aspects of the marine environment. The Seacoast Science Center preferred it to be a stereoscopic computer display with a fly-thru interface and wanted the main content to 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.

II. BACKGROUND

There are many areas of prior research relevant to issues dealing with 3D virtual kiosks. Some of these include cognitive theories 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; and studies of whether active learning environments are better than passive learning environments. It is also important to look at virtual museum environments that are currently in use. A discussion of these is in the following sections.

A. 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). 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 (Hitch, 1987).

Central to modern cognitive theory is the concept of working memory. Working 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 (a phonological loop). Some theorists also propose an additional executive buffer storing instructions on operations to execute. 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. The kind of information processed (visual or verbal) determines where it is stored (in the sketchpad 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, Paivio, 1991). 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.

B. Multimedia theory

Multimedia theory uses dual coding theory as a foundation (Mayer, Anderson, 1992). The central claim is that presenting information using more that one sensory modality will result in better learning. For example, if a student sees a picture of a dog with the label "dog" below it the student will process the picture in the visual channel and temporarily store it in visual working memory. 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 connect the picture of the dog and the label “dog” which will strengthen the encoding between them. A picture with words excites both the verbal and the visual processing systems whereas spoken (or written) words alone only excite the verbal system. The belief is 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, 1997).

Mayer and Moreno (1998) 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. This has become known as the SOI (Select, Organize, and Integrate) model. 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 connections between a proposition and its corresponding image.

C. Linking images and words

In human-to-human communications, a common way that people link what they are saying to something in the local environment is through a 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. Speech and gestures, such as pointing, are generally synchronized in time (Kranstedt, Kuhnlein, Wachsmuth, 2003) tending to occur at the beginning of an expression (Oviatt, DeAngeli, Kuhn, 1997).

Connecting images with audio through deixis, while navigating, is the function of some virtual pedagogical agents such as Cosmo (Johnson, Rickel, Lester, 2000). Johnson, Rickel, and Lester (2000) define 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.

1) Common audio activation methods

Three common audio activation methods are common in virtual environments. Direct selection (Hanke, 2002; Barbieri, Garzotto, Beltrame, Ceresoli, Gritti, Misani, 2001; Ressler, Wang, 1998) proximity (Ressler, Wang, 1998; Stock, Zancanaro, 2002; Guide-Man, 2002), and navigation based (Ressler, Wang, 1998). In direct selection, the user switches from navigation mode to selection mode and in this state clicking on an object activates the audio associated with it. This is a well-known feature in adventure-style computer games. Both real and virtual museums use the proximity method. In this method coming close to an artifact triggers the sound associated with it. For example, in a museum in Trento, Italy (Stock, Zancanaro, 2002), when a visitor wearing an audio headset comes close to a particular artifact, the audio begins to play. In the computer-based navigation method, when someone navigates through a virtual doorway, for example to enter a room, audio associated with that room begins to play (Ressler, Wang, 1998). This method is also quite common in video games.

D. Active Learning Versus Passive Learning

It is widely held that the activity of making connections between images and words is what aids in learning. As a Chinese proverb (attributed to Confucius) states: *"Tell me, I forget. Show me, I remember. Involve me, I understand"*.

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, Moreno, Boire, Vagge, 1999; Duffy, Cunningham, 1996). Constructivism is the idea that individuals construct knowledge based on their own personal experiences and not by simply acquiring it. The need for active involvement is also central to situated learning theory (Lave, Wenger, 1991; Brown, Collins, Duguid, 1996) and engagement theory (Kearsley, Shneiderman, 1998). Empirical studies involving children supports the idea 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).

Researchers have studied the benefits of using Virtual Reality (VR) from a constructivist perspective, and claim that VR provides precise and instinctive interaction with data (Bricken, Byrne, 1993), thus stressing the importance of using VR within a constructivist framework. Other findings also support the idea that 3D virtual worlds can be learning environments (Bricken, Byrne, 1993; Kelly, 1993).

Activity theory intends to provide a framework for building constructivist-learning environments (Johassen, Rohrer-Murphy, 1999). It proposes that activity is a precursor to learning instead of traditional views of learning before doing. Acting builds knowledge, building knowledge causes an adjustment in actions, which in turn changes knowledge, and so on. In other words, there mutual feedback exists between knowledge and activity (Fishbein, Echart, Lauver, Van Leeuwen, Langmeyer, 1990). The assumption is that interaction helps 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).

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, 1999; Melanson, Kelso, Bowman, 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, users will pay more attention to learning the controls and responses of the environment instead of learning the content. This might negate the positive benefits of active exploration.

III. EXHIBIT REQUIREMENTS AND DESIGN

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

The exhibition software, GeoExhibit, developed on the foundation of a visualization system for viewing geographic data spaces with a zooming user interface, called GeoZui3D (Ware, Plumlee, Aresenault, Mayer, Smith, House, 2001). The system is capable of displaying the following objects:

- A digital terrain model of the seabed (Bathymetry):
- Images shown as billboards in the 3D environment. These can have text labels
- Movie clips also shown as billboard
- A 3D dive sled that acts as a proxy for the user. It lies ahead of the user view, 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.
- Audio clips.

The 3D scene sits within a 3D data representation of the bottom of the Piscataqua River (approximately 800 meters across). A color-coded chart, displayed on the lower left of the user's screen, represents different levels of depth in the river (Fig 1).

Movie clips and images sit throughout the length of the river at various locations. Each image or movie clip has a label below it. Figure 1 shows the environment from the visitor's perspective.

Also associated with each image and movie clip is an audio clip that gives any where from 5 to 28 seconds of information. In Figure 2, an overhead view of the layout of the virtual environment shows an "x" at each location where there is an image with an associated sound clip.

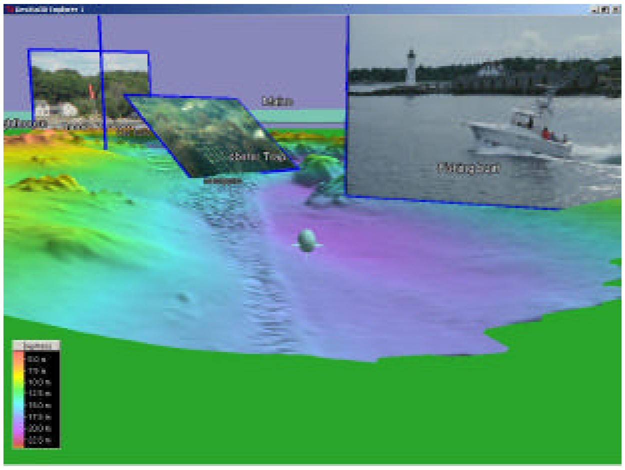

Figure 1. Users' view of the virtual Piscataqua River.

The user can freely navigate to 14 images (with associated sound clips).

The digital terrain model delineates the boundary of this virtual environment. If the user drives into it going down or to either side the vehicle will bounce 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. This acts as a kind of proxy for the user within the scene, although it is ahead of the user's viewpoint to keep it in the field of view (see Fig 1).

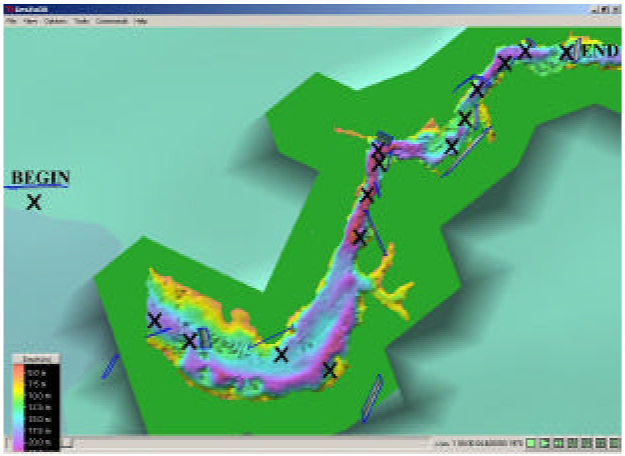

Figure 2. Overhead view of the virtual Piscataqua River

The vehicle moves by using the yoke input control illustrated in Figure 3. This is a rugged device, designed for use by the arcade video game industry. It can rotate like a steering wheel as illustrated and tilt..

Pushing the device forward against a spring force controls the forward velocity. Rotating the device controls the rate of turn. Tilting the device allows the user to move up, down or horizontally. Two trigger buttons at the front of the steering device (one for each hand) choose menu selections, and possibly "shooting" images in the tour to activate their corresponding audio descriptions.

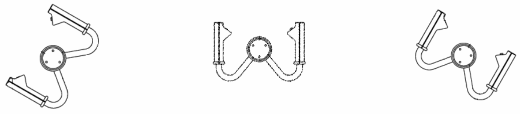

Figure 3. Input controller

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". A menu appears to ask the user to select a tour type. This menu is an introduction to selection in a non-navigational mode using the input controller. To make a selection the controller turns to the right or the left and while it remains in that position, a button is pressed. After selecting the self-guided tour (the right turn choice) the user's viewpoint smoothly moves to a designated start position 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 scene and receives a brief tutorial explaining how they should navigate. At this point, the input device activates and the user can freely navigate using the control yoke.

IV. AUDIO ACTIVATION METHODS

Iterations on the design of the exhibit began before any evaluative testing took place on the audio activation methods. As part of this process five distinctive audio activation methods were developed with the objective of finding one that would encourage interaction with objects in the exhibit, allowing audio information to be associated with the appropriate visual objects. The research literature, educational theory, and an interest in the difference between active and passive modes of interaction influenced the development of the methods.

Each of the five methods builds upon the characteristics of the one before it like a pyramid. There are trade-offs between the simplicity of the lower levels of the pyramid and the higher interactive upper levels of the pyramid. Simpler interfaces use less cognitive resources to learn how to navigate/activate audio-clips than with highly interactive, more complex methods of navigation/activation of audio-clips. Higher interactive interfaces could deter non-computer game players from trying it out, while simpler interfaces could bore computer game players and deter them from continuing to use it. Prior to the study it was not clear which of the five would be the best due to these trade-offs.

A. Proximity within a Zone

The first method "Proximity within a Zone" is the simplest. 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 solid angle out from the center of the image – Fig. 4).

The size of the zone is proportional to the size of the image, since smaller images become visible better at a closer distance and larger images from farther away. An anticipated problem with this design was that the activation zone was not visually evident 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.

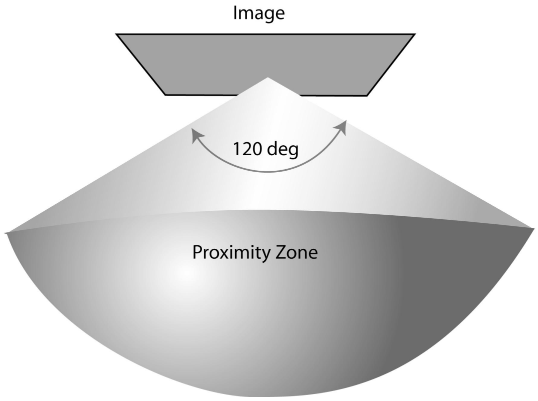

Figure 4. Proximity audio activation method

B. Visual Zone Cue

The second audio clip activation method added a visual cue to help the user associate the audio playing with its image. The visual cue is 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 synchronized with the playing of the audio clip associated with the image. It was the hope that adding the visual cue 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 activates, the visual line shows which way to turn to see the associated image.

C. Heading

The third method required the user to be facing an image for the audio to activation. It follows social behaviors of communication, people face each other in order to initiate communication. Audio begins to play when the center of an image is within 17.5 degrees of the forward direction and the vehicle is within the proximity zone.

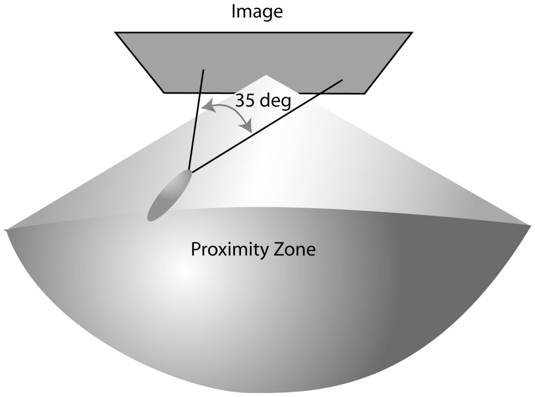

Figure 5. Heading audio activation method

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.

D. Button Press

The fourth method adds the requirement that when in the proximity zone the user press a button on the control yoke to activate audio. Adding the button press allowed the evaluation of a more dynamic involvement in activating each sound clip. A thick yellow frame appears around the image (normally all images are outlined in blue) when the user was in a position to turn on the audio on by pressing a button. Once the audio begins to play, the blue line visual cue appears then disappears when the vehicle has left the activation zone.

This method requires additional instruction at the start of the exhibit; teaching users how to use the buttons, as well as navigate with the steering device.

E. Tractor-beam

The fifth method has 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. This optimum viewing position being a position that is near the center of the image and at a distance from the image where it fits into the user's frame of view. The tractor-beam engages when the pressing the button within the activation zone facing the image (again the yellow highlighted frame visual cue appears). After the button is pressed, the user temporarily loses control of the vehicle and the vehicle smoothly translates and rotates to the optimum viewing position. This repositioning typically takes under 1 second. When the vehicle arrives at the optimum viewing location, the sound clip is activated (the blue line visual cue appears on audio activation then disappear when the vehicle leaves the activation zone). Control of the vehicle returns to the user after the tractor-beam animation is complete enabling the user to continue navigation at any time.

An anticipated problem was possible confusion with loss of control as the tractor-beam activated.

V. EVALUATION OF AUDIO ACTIVATION METHODS

It was the intent that each of the five audio activation methods was a plausible best solution to the problem of linking verbal explanation with imagery in a 3D navigational environment. These methods lie on a continuum from least to most interactive. In particular, the last two methods required active button presses to activate the audio material.

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 plays. Earlier predictions claimed 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. 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.

A. Study 1: Testing with Museum Visitors

Unlike most human factors evaluations, where informing people they are in an experiment is common, this evaluation of the five methods for activating audio sound clips this took place in a natural museum environment. The visitors learned only upon request, that the exhibit was under observation (with no mention of the evaluations that occurred, so as not to influence their natural behavior).

1) Subjects

The kiosk subjects were visitors of the SSC in Rye NH, and the New England

Aquarium (NEAq) in Boston. There were 100 subjects in total. 20 different users tried each of the 5 different audio clip activations methods. The breakdown of the subjects is in Table 1.

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

T1				T2				T3				T4				T5			
20				20				20				20				20			
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 goal was to obtain 20 subjects per audio clip activation method. This came about by changing the method every few subjects. Serendipitously, there were roughly an equal number of adults as were children, this may have been due partly to the fact that there were mostly children at the SSC and mostly adults at the NEAq. The children's age group was made up of children from age 5 thru 8th grade and the adult age group was made up of 9th graders and older.

2) Procedure

As visitors approached the exhibit, they learned (if asked) that the exhibit ran on a trial basis with the public. (At the SSC, 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 NEAq, a more portable exhibit was set up near the entrance of the aquarium. With this was also necessary to inform the patrons that it was indeed available for use.

3) *Measures*

The program controlling the exhibit gathered data on the time facing each picture and the number of activation zones entered.

When each subject finished their exploration, they had a choice to participate in a brief exit interview. The exit interview contained questions about various parts of the exhibit. The design of two of these in particular was specifically to evaluate the effectiveness of the visual cues; the other questions related to user interface and preferences and do not need to be discussed. The two relevant questions were: 1. Did you know what the blue line (visual cue) coming from the vehicle was for? 2. Did you notice the yellow frame highlight around the image (if using a button method)?

B. *Results of Study 1*

The average time a user was oriented toward an image in a zone with audio activated was a measure of the user's level of interest (Stock, Zancanaro, 2002). This captured the fact that the user was probably both "looking at" an image and "listening to" the audio associated with that image. A count of the number of activation zones entered by the user was a measure of the overall ease of navigation. Separate analyses of variance ran with each measure as a dependent variable.

Figure 6 and 7 summarize the time-oriented-toward-images results. The ANOVA revealed 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 showed 2 groups of audio activation methods. The tractor-beam (T5) and button press (T4) methods made up the first group (active methods) with the longest average times [8.2 seconds]. In the second group were the zone (T1), visual cue (T2) and heading (T3) methods (passive methods) with an average time of 2.2 seconds. An age-method interaction [F(4,80) = 2.52, p < .05] also became evident. Adults spent more time (9.4 sec) than children did (3.2 sec) facing images when using the button press (T4) method. Both adults and children faced images (with audio playing) about the same amount when the tractor-beam (T5) method is used (10.4 and 9.5 seconds on average, respectively).

There was also gender-method interaction illustrated in Figure 7. [F(4,80) = 5.68, p < .0001]. This reflects that males faced audio activated images longer while using the tractor-beam (T5) method, whereas females had greater times for the button press (T4) method. It is possible that this interaction is spurious due to the small number of female visitors using the tractor-beam (T5) method (only 4 females total, with 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 one was a 3rd grader.)

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].

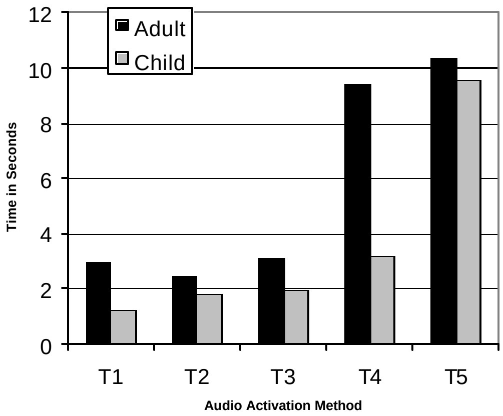

Figure 6. Average time facing images per activated audio clip according to age per method used.

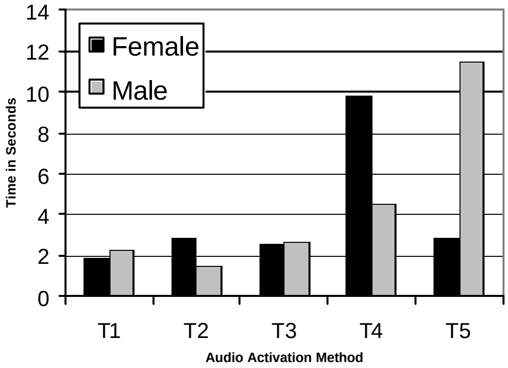

Figure 7. Average time facing images per activated audio clip according to gender per method used.

1) Exit Interview Results

The data recorded for the exit interviews is incomplete due to fact that many of the participants declined participation. In addition, of those that did participate not all of the people answered all of the questions.

Table 2 shows the “yes” answers to two questions, 1. “Did you know what the blue line (visual cue) coming from the vehicle was for?" 2. "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 section 4). In all methods but the zone method question 1 was asked, only the button press, and the tractor-beam methods received question 2.

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

	Question 1	Question 2
T1	N/A	N/A
T2	12/14	N/A
T3	9/10	N/A
T4	10/13	7/14
T5	7/13	7/14
ALL	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 that is was difficult to see yellow highlighted frame.

C. Discussion of Study 1

The main result was that users spent more than three times longer in front of the images when using 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. In addition, adults on average spent more time facing images than did children, perhaps because children were more interested in driving around the 3D environment than in listening to the audio content.

The age-method interaction indicates that adults and children react differently according to the method of audio activation they are using. In particular, adults, using the button press (T4) method, spent three times as much time facing images with activated audio as children. With the tractor-beam method, children may have not realized that they could move off once they were in front of an image.

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. One possible explanation for these results is that males, like children, were not aware of the return of control so they lingered longer.

The finding that that the number of zones entered did not vary significantly suggests that none of the methods affected user's ability to navigate through the environment.

D. Study 2: Subjective Comparative Assessment

The second method used to evaluate the audio-activation methods utilized a semi-structured interviewing technique. The goal of this was to obtain opinions from a group of interested adults who each experienced all five of the audio activation methods.

1) Subjects

Ten adult subjects (six female, four male) were solicited their help in evaluating the exhibit. Four of these were employees of the SSC, but not directly involved in the exhibit. The other six were visitors to the NEAq.

2) Procedure

Each participant had an opportunity to try all five audio activation methods, with a different random order for each subject. Following each method, subjects heard the same set of questions that were in the exit interview for study 1. When subjects had tried all five methods, they ranked them in order of preference.

The average time for this protocol was approximately 15 – 25 minutes per person.

3) Measure

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

E. Results and Discussion of Study 2

The average rankings for each audio activation method used are in Figure 8.

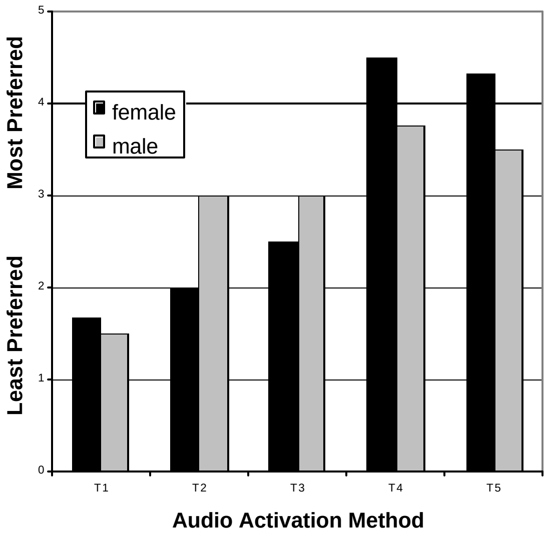

Figure 8. Mean ranking for audio activation method

The button press (T4) and tractor-beam (T5) methods obtained the highest mean rankings. Some of the reasons that were given for liking the button press method included, “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 received the following comments; “it positioned you for good viewing” and “[it] is good to be actively involved”. However, it was inferred, from observation and recorded comments that 4 of the 10 subjects did not realize they had lost control of the vehicle during the tractor-beam repositioning; hence, they did not realize there was a difference between the button press (T4) and the tractor-beam (T5) methods. Nevertheless, when told of the difference between them they liked the idea of the tractor-beam (T5) method better. The finding that both the button press (T4) and the tractor-beam (T5) were the top ranking methods 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 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.

VI. CONCLUSION

The results of both the objective and subjective phases of testing indicated that audio activation methods that involve an explicit act of selection (a button click) were superior to the methods where activation occurred by navigating into a particular location in front of an image. The button press (T4) and the tractor-beam (T5) audio activation methods yielded longer times facing the images with audio playing and received the highest ratings.

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. 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. The active methods used the less effective yellow border, despite which they still performed the best. This suggests that the active methods could use improvement.

As noted in our introduction, there are trade-offs between the less active and highly interactive methods of audio activation. 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 she cannot 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. At the same time, the active activation of audio may encourage the cognitive binding between the audio and the text. On the other hand, the active methods may also have been harder to learn. For many of the older visitors learning to navigate appeared to place them at the limits of their tolerance for new technology. Having to learn that pressing the button was necessary to activate audio added to the learning requirement. Nevertheless, the overall preference of active methods over the more passive methods of audio activation supports constructivist theories.

One of our goals in this research was to develop a method suitable for use the SSC Exhibit. As the results turned out, the button press and tractor-beam were markedly better than the rest. From the empirical results, the button press method appeared not to be as effective with children whereas the tractor-beam method appeared to be not as effective with female visitors. Nevertheless, females ranked the tractor-beam method highly.

Our final decision was to adopt the tractor-beam method. 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. It appeared that some of the shortcomings of the tractor-beam method would improve with further development arising from comments of the subject. Adding an audible hum and simultaneously showing the blue line from the vehicle proxy to the center of the image during the repositioning process addressed the problem of feeling of a loss of control.. This made it clear that factors other than the user's input were causing the movement. The last update was placing a yellow "+" in the middle of an image when subjects were within activation range to make the zone cue more visible.

The novel tractor-beam method of audio clip activation proved to be arguably the best of the five implemented. In its final form it works as follows: the tractor-beam activates 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 she presses the trigger button. At this point, a ray (blue line) links the avatar with the center of the image and the user temporarily loses control of the avatar while it smoothly repositions 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 returns back to the user. The properties of the tractor-beam causes 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 SSC.

ACKNOWLEDGMENT

The authors would like to thank Tracy Fredericks at the SSC for the effort she put into this project and her willingness to work with us at any moment. The help of Roland Arsenault in building the hardware interface was also much appreciated. The authors are grateful for funding received through a grant from the SSC and NOAA Grant NA170G2285.

REFERENCES

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

Barbieri, T., Garzotto, F., Beltrame, G., Ceresoli, L., Gritti, M., & Misani, D. (2001). From dust to stardust: a Collaborative 3D Virtual Museum of Computer Science in *Proceedings ICHIM 01*, Milano, Italy, 341 – 345.

Bricken, M., & Byrne, C. (1993). Summer students in virtual reality: a pilot study on educational applications of virtual reality technology. In A. Wexelblat (Ed.), *Virtual Reality: Applications and Explorations*. (pp. 199 – 217). San Diego, CA: Academic Press.

Brooks, F.P. (1988). Grasping Reality Through Illusion: Interactive Graphics Serving Science. *Proceedings of the Fifth Conference on Computers and Human Interaction*, ACM, 1-11.

Brown, J.S., Collins, A., & Duguid, S. (1989). Situated cognition and culture of learning. Educational Researcher.18(1), 32-42.

Clark, J.M., & Paivio, A. (1991). Dual Coding Theory and Education. *Educational Psychology Review* 3(3), 149-170.

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

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

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

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

Fishbein, H.D., Echart, T., Lauver, E., Van Leeuwen, R., & Langmeyer, D. (1990). Learners' Questions and Comprehension in a Tutoring Setting. *Journal of Educational Psychology* 82(1), 163-170.

Guide-Man (2002). Audio Guides, Ophrys Systems [On-line],1-10. Available: http://www.ophrys.net/audioguide%20english/documentation/GM-angl.PDF

Hanke, M.A. (2003). Explore the Fort at Mashantucket, Design Division, Inc., of New York [On-line]. Available: http://www.pequotmuseum.org/Home/AboutTheExhibits/InteractiveExhibits.htm#

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

Hibbard, W., & Santek, D. (1989). Interactivity is the key. *Proceedings of the Chapel Hill Workshop on Volume Visualization*, 39 – 43.

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.

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

Jonassen, D., & Rohrer-Murphy, L. (1999). Activity Theory as a Framework for Designing Constructivist Learning Environments. *Educational Technology Research and Development* 47(1), 62-79.

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

Kelly, R.V., Jr. (1994). VR and the educational frontier. *Virtual Reality Special Report*, 1(3), 7-16.

Kranstedt, A., Kühnlein, P., & Wachsmuth, I. (2003). Deixis in Multimodal Human Computer Interaction: An Interdisciplinary Approach. University of Bielefeld, Germany, *Gesture Workshop*, Genova, Italy, Springer-Verlag, 112-123.

Lave, J., & Wenger, E. (1991). Situated Learning: Legitimate peripheral participation. Cambridge, UK: Cambridge University Press.

Travis, D., Watson, T., and Atyeo, M. (1994). Human psychology in virtual environments. In L. MacDonald and J. Vince (Eds.), *Interacting with virtual environments*. (pp. 43-59). Chichester, UK: John Wiley & Sons.

Mayer, R.E., & Anderson, R.B. (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., Moreno, R., Boire, & Vagge. (1999). Maximizing constructivist learning from multimedia communications by minimizing cognitive load. *Journal of Educational Psychology* 91(4), 638-643.

Mayer, R. E. & Moreno, R. (1998). A Cognitive Theory of Multimedia Learning: Implications for Design Principles. Paper presented at the annual meeting of the ACM SIGCHI Conference on Human Factors in Computing Systems. Los Angeles, CA. [On-line]. Available: http://www.unm.edu/~moreno/PDFS/chi.pdf

Mayer, R.E., & Moreno, R. (1998). A split-attention effect in multimedia learning: Evidence for dual processing systems in working memory. *Journal of Educational Psychology* 90(2), 312-320.

Mayer, R.E., & Sims, V.K. (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.

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

Oviatt, S., DeAngeli, A., & Kuhn, K. (1997). Integration and Synchronization of Input Modes during Multimodal Human-Computer Interaction. in Proceedings of CHI 97, Atlanta, GA, ACM Press, 415-422.

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

Ressler, S., & Wang, Q. (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., & Allison-Bunnell, S. (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.

Stock, O., & Zancanaro, M. (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.

Ware, C., Plumlee, M., Arsenault, R., Mayer, L.A., Smith, S., & House, D. (2001). GeoZui3D: Data Fusion for Interpreting Oceanographic Data, *Proceedings Oceans 2001* 3, 1960 – 1964.

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.

Wilson, P.N., Foreman, N., Gillett, R., & Stanton, D. (1997). Active Versus Passive Processing of Spatial Information in a Computer-Simulated Environment. *Ecological Psychology* 9(3), 207-222.