2011

Interface design to support situation awareness in virtual puppetry

Keisha Joy Harthoorn
University of Northern Iowa

Interface Design to Support Situation Awareness
in Virtual Puppetry

A Thesis

Submitted

in Partial Fulfillment

of the Requirements for the Designation

University Honors with Distinction

Keisha Joy Harthoorn

University of Northern Iowa

May 2011

Introduction and Motivation

"Virtual Heritage" is the use of digital media to reconstruct cultures and cultural artifacts as they are today or as they might have been in the past. The central element is usually a three-dimensional computer model of a person, place, or thing. Frequently, these are ancient monuments, temples, homes, and other social spaces (Jacobson, 2008). The goal of Virtual Heritage is to draw viewers into the virtual world and allow them to directly experience the overall context of the environment. This phenomenon is known to researchers as "presence." It is a long held belief that the increased presence yields better the opportunities for deeper learning (Devine, 2007).

To increase presence, a lot of effort goes into the modeling and rendering of these environments to make them visually accurate (and believable); however, the illusion of presence in an alternate reality is fragile and easily disrupted. Virtual spaces are almost always empty, silent, and lonely. Lacking people, they lose social context and much of their meaning (Ulicny & Thalmann, 2002). In the same way that character actors bring life to living museums such as Colonial Williamsburg, this project seeks to bring to life the virtual environments with virtual people and provide a roadmap to populating other virtual heritage sites.

Technologies are readily available for adding human models to these virtual heritage sites. Autonomous agents can represent ancient peoples conducting their daily business or performing other orchestrated activities. It is even possible to have these virtual humans, powered by artificial intelligence (AI), interact with visitors to the environment. The problem arises from the fact that interactions based on current AI technology are fairly narrow and often somewhat scripted (Mateas, 2001). Robust interaction between the audience and virtual characters can still benefit human interpretation of audience questions, reactions, and gestures.

This is not trivial; when real people interact with one another, they have the advantage of being able to tap into a wealth of deictic behaviors and gestures that are practically involuntary. In addition, there are social constraints that go along with interacting with the audience (making eye contact, for example). These are second nature for human presenters, but not easily replicated by computer generated characters.

The broad objective of this research is to develop and evaluate techniques to allow actors to manipulate characters in virtual heritage environments that are capable of rich and engaging interaction with the audience. The interface designed should enable the puppeteer to quickly and easily interact with both the virtual world and their audience, while at the same time giving the audience an experience that is more advantageous for learning. The success of these virtual characters requires substantial attention to the user interface for the human operator. Not only must the virtual character support various gestures and behaviors, but they must also be easily and effectively toggled by the operator. This requires that the operator be aware of events that are taking place in both the environment and in the audience and be able to respond to each appropriately.

Design Principles

In studying this problem, it was determined that the system should adhere to the following design principles:

- Cost Effective: The interface design should use relatively inexpensive materials and software. While there may be a few museums or learning centers that can afford multi-million dollar setups, most will be operating on a limited budget. It would be most beneficial if this interface were usable in a wide variety of virtual heritage sites. This means using commercial, off-the-shelf components whenever possible.

- Limited Training**:** Little experience should be needed in order to be able to work with the system. Again, it is unlikely that many places will be able to afford hiring someone who is specifically trained to operate a certain interface. Most likely, staff will be limited and may even be volunteers. Thus, a system that requires hours of training is not feasible.
- Blended Control: Complete, direct control of the puppet may not be practical. Technologies exist that allow a person to put on a bodysuit and generate a virtual character that moves exactly as he/she does. This, however, requires that the operator focus more on acting and performance than on increasing the audience's learning. In order to avoid this, this research intends to keep the allowable movements as simple as possible without making the character seem scripted or robotic. The puppeteer will be allowed to select from a library of animation scripts to perform common gestures, e.g. random eye contact when addressing the audience as a whole. Pointing and gaze will be included as well.
- Multiple Perspectives: Because the operator is required to interact with the live audience and the virtual environment, multiple perspectives are needed. He or she will need to be able to see what is going on in the virtual world while simultaneously maintaining a view of the real world. In order to smoothly interact with both, the puppeteer needs to be able to transition his or her attention between these views almost instantaneously.

In order to adhere to these design principles, it is critical for the interface to aggressively support situation awareness. Endsley (1995), who is recognized as a world leader in studies of situational awareness in advanced systems, defined situation awareness as: "the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future" (p. 5). The operator of a virtual heritage interface must understand events that are taking place in both the environment and in the audience in order to respond to each appropriately. It is likely that environmental awareness can be reinforced by training; the configuration of environmental features may be learned over time. Reaction to the audience, on the other hand, requires the operator to be aware of a unique group of spectators with each presentation. For example, how might the actor/avatar acknowledge which audience member wishes to ask a question?

It is assumed that the integration of the guide with the environment has the potential to increase an audience's sense of presence in a virtual heritage environment. To validate this theory, the impact of a virtual character needs to be measured against the experiences with a live docent. This requires the development of a set of metrics that can be used to assess audience engagement and learning. Incrementally refining the character to exceed the docent will reveal attributes of the intermediary interaction model that are essential to realize this potential.

Purpose

The purpose of this research is to develop an interface that allows an actor/puppeteer to control an avatar for a virtual heritage site. This interface should strive to support situation awareness, but not so much that the virtual character appears scripted or robotic. The project is focused on designing an interface that integrates with the work of Jeffrey Jacobson from the Carnegie Museum with his virtual reconstruction of an ancient Egyptian temple; however, these research findings should be applicable across multiple domains. The main goal of this research is to identify which form of interface will be most advantageous for virtual heritage sites. To this end, test subjects were asked to engage in simple task performance studies on different possible selection methods. The results of these studies will be used in determining how the most effective interface ought to be developed.

Literature Review

Numerous issues arise when considering the development of an interface that supports situation awareness to a sufficient degree in virtual heritage sites. Such issues include: determining the effectiveness of direct control against that of indirect control, comparing and contrasting zero-order and first-order devices, the advantages and/or disadvantages of using a Wiimote®, how to best improve situation awareness, finding the pros and cons of large screen verses small screen displays, deciding how to best design an environment that will truly create a better learning experience for the audience, and finding a method to evaluate the effectiveness of the interface design. Previous research can be used to help determine the best way to deal with these issues.

The first issue to address is whether direct control (the movement of the body is the input to the display) is better or worse than indirect control (the movement of some peripheral is translated to movement on the screen). This research seeks to address this issue specifically by using Wiimotes® as direct control devices, and a mouse as an indirect control device. Determining which method, if any, creates the most effective environment for the system operator is an essential step with this project.

Forlines and Balakrishnan (2008) conducted a study meant to address this topic. They conducted two studies, one of which analyzed both direct and indirect pointing with a stylus device on a tablet PC. This study had twenty-four participants, ranging from ages 18 to 37. None of the participants had significant experience with a tablet PC or similar device. For direct input, the user's movements were tracked using the display on the tablet; the results of such movements were displayed in the same area. Indirect input was conducted in a similar manner, but the results of the user's movements were displayed on an external monitor instead of the tablet screen. In each test, participants were asked to point and tap back and forth between given targets on the screen.

The results of their study indicated that direct input tasks could be performed with greater efficiency than those of indirect input. The mean selection time between targets for direct input was significantly lower than that of indirect input; however, they also found that indirect input was more effective for the selection of smaller (and thus more difficult) targets (Forlines & Balakrishnan, 2008). These results indicate that the interface developed in this project would benefit from a direct pointing device (Wiimote®) instead of an indirect pointing device (mouse), but only as long as the targets are not so difficult that they counteract the benefits of using direct pointing.

Pointing tasks can also be defined in terms of first-order (control over the directional velocity of the cursor) and zero-order (control of the cursor is absolute). These are often related to Fitts' Law. P. M. Fitts (1954) proposed that the time it takes to rapidly move to a target is directly related to the distance traveled to reach that target and the size of the target area (p. 5). Campbell, O'Brien, Byrne, & Bachman (2008) used this law in their study of zero versus first-order control with a Wiimote® device. They claimed that: "A mouse may not be enough to guarantee easy use if the interface occurs in more dimensions than the pointing device is capable of manipulating," and thus wished to test a device they felt more capable of handling three-dimensional environments, the Wiimote®.

The study Campbell et al. (2008) conducted included thirty-nine participants with ages ranging from 18 to 40. Twenty-four of these participants had used a Wiimote® prior to the study, and all thirty-nine had normal or corrected normal vision. The researchers believed that none of the demographics in this population had a significant effect on the results. The study used custom Java software in order to create the graphical interface for the tests. These tests consisted of a black target moving back and forth between two red boxes of varying shapes to indicate which box the participants were to select. Participants were randomly assigned to a zero-order or a first-order group and were asked to complete fifteen practice trials and fifty recorded trials.

The results of the Campbell et al. (2008) study showed that participants using the Wiimote® as a first-order device took longer to move from point to point. The researchers claimed: "This would indicate that when using a free moving input device, it is still prudent to allow the user to control position rather than velocity on the device" (Campbell et al., 2008, p. 5). Due to the results of their study, the interface designed in this project shall use a first-order control device. The researchers went on to conclude: "In real applications for long-amplitude movements, the mouse is not always ideal, as it must occasionally be lifted and re-positioned mid-move (e.g., when making a long movements and bumping into the keyboard). The Wiimote® has no such limitation" (Campbell et al., 2008, p. 5). For this reason, and the reasons stated above, this research project will consider using the Wiimote® as a direct control, zero-order device. Although prior research shows that this is most likely the best controller configuration for the system, this project will still use its own studies in order to determine if it remains true for this specific interface design.

The matter of how to best improve situation awareness must also be addressed in this project. According to Endsley (1995), situation awareness (SA) is a state of knowledge obtained from the processes used to reach that state. Put in layman's terms, it is simply defined as "knowing what is going on." There are three levels of SA. Level 1 SA is the perception of elements in the environment. This describes a being's knowledge of surrounding objects and the important attributes of those objects. Level 2 SA is the comprehension of the current situation. This level is achieved by combining knowledge obtained from separate elements in Level 1 SA. It establishes the significance of Level 1 elements related to the operator's goals. Level 3 SA is the projection of future status. It is the ability of the operator to determine the future states of surrounding objects (at least for a short period). It should be noted that these levels are not always mutually exclusive. In addition, situation awareness at any level is not instantaneous; it is established over a period of time (Endsley, 1995).

Situation awareness can also be linked to performance (although not always directly). Typically, poor performance ensues when SA is inadequate for the system operator's needs. This is not always the case, however. Ample performance can be maintained if the operator understands his/her insufficient level of SA and is able to adapt accordingly. Endsley claimed: "Good SA can therefore be viewed as a factor that will increase the probability of good performance but not necessarily guarantee it" (Endsley, 1995, p. 9).

Endsley (1995) went on in her article to discuss how SA impacts interface design. She stated: "The way in which information is presented via the operator interface will largely influence SA by determining how much information can be acquired, how accurately it can be acquired, and to what degree it is compatible with the operator's SA needs" (p. 19). Generally, it is best to develop a design that will meet all of the operator's needs without causing them undue cognitive effort. The amount of SA provided needs to be balanced against the operator's mental workload. Endsley (1995) explained eight methods for improving SA with interface design, listed below:

> 1. Displays should show a greater degree of information that is processed in terms of Levels 2 and 3 SA.
> 2. The interface should provide information that is directly related to the system operator's goals.
> 3. The interface needs to be designed such that cues indicating the presence of an atypical situation are prominently displayed.
> 4. The interface design should take both top-down (goals and expectations influence how attention is directed) and bottom-up processing (cues activate goals and models) into account.
> 5. Global SA ought to be provided to the operator. This means that the operator has an overview of his/her goals while simultaneously being provided all relevant information for his/her immediate needs.
> 6. Displays should be designed to filter out extraneous data that may distract the operator from his/her goals.
> 7. When applicable, the interface should provide information on future states of the system in order to improve Level 3 SA.
> 8. Systems should support the parallel processing of information. (p. 20)

This project will take all of these design suggestions into account in order to produce an interface that is capable of producing the highest level of SA possible.

As mentioned earlier, another issue that arises while making considerations for this research project is whether or not large screen displays can significantly affect performance. Czerwinski et al. (2003) sought to answer such an issue in their study of large display interfaces. In the study, they compared a user's performance using a 15" display monitor against that of a 46.5" display monitor. Those studied were asked to perform a series of web and Microsoft Office task steps on a Windows XP operating system. There were fifteen participants, seven of which were females and eight of which were males, with ages ranging from 23 to 50. These participants were evaluated as intermediate to expert Windows and Microsoft Office users, and each of them had normal or corrected normal vision. None of the participants reported using multiple monitors at home prior to participating in the study. Czerwinski et al. (2003) tentatively hypothesized that the users would be more effective at accomplishing their tasks on the larger surface because of reduced window management.

Each participant was asked to conduct twelve tasks per display, two of which counted as trial runs and not assessed in the final results. Every task had the same eight-step sequence, which included:

> 1. A phone number was presented which had to be remembered throughout the trial.
> 2. A web page target (title and summary description) was presented to the user upon removal of the phone number, and the user was to come up with 3 search terms for searching for this target before continuing.
> 3. Alta Vista's search page was presented and the participant was to type in the 3 search terms.
> 4. The best match from the search list was to be selected.
> 5. The participant was to determine who designed the web page (a computer science student, a small software company or an upscale design firm).
> 6. The URL from the web page was to be copied and pasted accordingly into a Word document containing the 3 design categories.
> 7. An image of the web page was to be captured (using Alt + PrintScreen) and pasted into an empty PowerPoint slide deck (empty slides were already prepared for this task).
> 8. The participant pressed a button to conclude the task trial, and then had to type in the phone number from memory. (Czerwinski et al., 2003, p. 5)

Each participant was given a five-minute time limit to complete these eight steps for each task. The total session lasted approximately two hours.

Czerwinski et al. (2003) found that phone number memory was higher, task times were significantly faster, and user satisfaction was greater with the large-screen display. There were some usability issues reported for each display type. For example, small screen display users would waste time during the completion of their tasks by sizing and repositioning windows. In addition, they often opened windows that they did not intend to because the task bar would aggregate windows by application. Large screen display users complained that the screen was too bright and they would have preferred to have been permitted to sit farther back from the display when interacting with it. The researchers concluded: "This study demonstrated that there is a significant performance advantage to using very large, multiple monitor display surfaces while carrying out complex, cognitively loaded productivity tasks on the computer" (Czerwinski et al., 2003, p. 6). Their findings suggest that use of a large display would be beneficial in the design of this interface.

This project also seeks to determine whether or not the learning experience of the audience will be improved by the use of a digital interface. Since the current design is geared towards virtual heritage sites in museum settings, this becomes an important issue. This project will focus on developing a design that will be beneficial for the system controller, but the experience of the audience must also be considered. According to C. Dede (2009), "Studies show that immersion in a digital environment can enhance education in at least three ways: multiple perspectives, situated learning, and transfer" (p. 1).

The ability to switch between multiple perspectives is a valuable way of comprehending a complex event. Most often, this is achieved by toggling between exocentric and egocentric frames of reference. An exocentric frame of reference provides an external view of the event, while an egocentric frame of reference provides an internal view. Each frame of reference benefits learning experiences in different ways:

> A major advantage of egocentric perspectives is that they enable participants' [actionable] immersion and motivation through embodied, concrete learning, while exocentric perspectives foster more abstract, symbolic insights gained from distancing oneself from the context (seeing the forest rather than the trees). Bicentric experiences that alternate these views combine these strengths (Dede, 2009, p. 4).

Situated learning is another method of learning improvement Dede (2009) mentions in his article. This type of learning requires real learning experiences gained through models, mentoring, or participation. Even though this tool has a lot of potential to increase learning, it is seldom used because it is less structured than traditional teaching methods. Interestingly, Dede found in his study that immersive interfaces could implement this teaching style effectively. "Our research results...show that a broader range of students gain substantial knowledge and skills in scientific inquiry through immersive simulation than through conventional instruction..." (p. 6).

Dede's third method of education improvement mentioned in his article is that of transfer. The term *transfer* refers to applying knowledge gained in one situation to another situation. It is demonstrated if instruction on a certain task leads to increased performance on a transfer task. A major criticism of today's instruction methods is the low rate of transfer they produce. Dede stated:

> Even students who excel in educational settings often are unable to apply what they have learned to similar real-world contexts. The potential advantage of immersive interfaces for situated learning is that their simulation of real-world problems and contexts means that students must attain only near transfer to achieve preparation for future learning (Dede, 2009, p. 8).

The conclusions reached in Dede's research indicate that the audience will indeed benefit from a digital immersive experience such as the one being developed for this research project.

This study also required a method to evaluate the effectiveness of the interface design. In 1954, P.M. Fitts conducted a study designed to test the relationship between the speed, amplitude (size), and accuracy in simple selection tasks. He conducted three separate experiments. The first experiment consisted of reciprocal tapping tasks, where participants were asked to tap back and forth between two rectangular metal plates with a stylus. Their task was to alternately hit the center of each plate as many times as they could. There were sixteen different conditions on which they were tested, with varying target sizes and distances between them. In this experiment, he found that, within each target distance tested, the movement time increased as the target size decreased. In the same way, within each target size, the movement time increased as the distance between the targets increased (Fitts, 1954).

The second experiment conducted by Fitts required participants to transfer washers from one pin to another. Again, the target size and distance between them varied. The results of this experiment were similar to those of the first. Fitts' third experiment asked participants to transfer pins from one set of holes to another. The results he found from this experiment coincided with those of the previous two. Based on his results, Fitts (1954) proposed a method for calculating the index of difficulty in similar selection tasks. His findings are widely supported, and his formula, commonly called Fitts' Law, will provide this research with a method for calculating the index of difficulty in selection tasks.

Such previous research can help to provide the basis for the hypotheses in this experiment. Forlines and Balakrishnan (2008) suggest that direct order control devices would be preferred in most interfaces, and Campbell et al. (2008) suggest that zero-order control devices are more beneficial. Endsley's (1995) work advocates simpler, more intuitively designed interfaces in order to increase situation awareness. Because of works such as these, hypotheses can be drawn about how to best design the proposed interface in a way that improves both performance and situation awareness.

Hypotheses

1. The interface that relies on the absolute (Wiimote®) controller will produce faster transition times than the relative (mouse) controller. The Wiimote® is a direct-order pointing device, which means that users do not have to map their actions from a device to the screen; the cursor is always directly related to where they are pointing. This is assumed to allow for faster reaction times and increase situation awareness by decreasing the cognitive load.
2. The "point to screen" technique will have minimal transition time. It should be almost instantaneous, provided that the operator can learn the mapping from the controller to the peripheral screen. (Note: A learning effect is anticipated here). However, it is likely that this transition will incur a significant disorientation to relocate the mouse position and the target. The “toggle” technique will take more time to transition – although this can be minimized. Such minimal disorientation is to be anticipated. The operator’s physical attention is focused on the peripheral screen during the transition, and he/she begins his/her efforts to acquire the target from a relatively fixed location. These interactions add to the time required to locate the next target, which then increases the overall transition time.

Methodology

Design

The interface designed for this study contains three smaller, peripheral displays, and one focal display. The three peripheral displays allow the puppeteer to remain aware of events in both the virtual environment and the audience, while still focusing on the larger main display.

When the operator wishes to interact with a display, he or she must first bring it into focus. This creates two problems. The first of these is: how do users switch between screens? And second: how can the selection of objects within a screen be controlled?

In order to answer these questions, this study focused on determining the best method for bringing peripheral displays into focus. Information that is contained on the peripheral displays periodically requires primary attention of the operator. This requires that the operator change focus based on peripheral information. Such a change is to be accomplished by swapping the information contained on the smaller peripheral display to the significantly larger primary display. Two techniques, toggle and "point to screen," were compared as methods for accomplishing this task.

Another focus of the study was to determine how to best select objects once they were brought into the focal display. Direct pointing is one such way of doing this. Direct pointing simply means that the cursor is located in direct relation to where a user is pointing. It was assumed that this would decrease cognitive load, and therefore increase situation awareness. Direct pointing methods also allow the puppeteer to be able to select elements in multiple displays without reorienting. As a result, this research seeks to use a device that could incorporate this form of control.

A device well known for its direct-pointing capabilities, the Wiimote®, was chosen for this purpose. In addition to being a direct-order device, it fits into this study's design principles because it is relatively inexpensive and requires little training to use. The Wiimote®'s performance was to be evaluated against that of the mouse, an indirect form of control. This comparison was made in order to determine if the Wiimote®'s direct pointing capabilities were indeed superior to a relative pointing alternative.

As mentioned earlier, this study was designed in order to determine the best way of creating an interface capable of maintaining a sufficient level of situational awareness. In order to determine which form of control would be most beneficial for the interface design, it was necessary to test whether there is a significant difference in the Fitts' curve between relative (mouse) and absolute (Wiimote®) controllers. The test developed required participants to select targets of varying size that were randomly generated on the screen. The time it took the participant to click the target and the distance between two clicks was used to determine the index of difficulty. The data taken from these trials was then compared to Fitts' Law (Fitts, 1954).

Another focus of the study was to determine the best method for bringing peripheral displays into focus. Information that is contained on the peripheral displays periodically requires primary attention of the operator. This requires that the operator change focus based on peripheral information. Switching a peripheral display to the focal display gives the human operator greater control over the result of his/her actions, and allows the operator to focus in on who or what is currently demanding attention. Thus, the user is required to engage in three distinct activities: 1) The user must identify that the information that they need is contained in the peripheral display. 2) The user must bring the peripheral display into focus. 3) The user must act on the information contained in the display.

The first activity is most likely a function of the conspicuity of the target and the configuration of the displays. The third activity is comprised of acquiring the target and can be predicted by Fitts' Law (Fitts, 1954). The second activity is influenced heavily by the interface design of the entire (focus + peripheral) display. Bringing the peripheral information into focus requires the user to engage in the physical act of issuing the command to swap screens and the cognitive act of reorienting themselves to the new display. Two techniques were compared for bringing the display into focus: 1) Toggle - A button press corresponding to the desired peripheral screen causes it to take focus. The location of the mouse pointer remains in the same relative screen location. 2) Point to Screen - The user gestures toward the peripheral screen that they wish to activate. The mouse pointer is repositioned to the relative entry/exit point of the focus screen. Participants were given a series of tests to calculate transition times between the displays. The results of this study are meant to aid in the decision on which control scheme is best for the puppet control interface.

The interface developed for this research will eventually use the Unity game development engine in order to generate the virtual puppet and its virtual environment. The puppeteer will be provided three peripheral views, two of which are to be driven by Unity. The third view will be generated by a webcam or video capture device and integrated with the rest of the interface via custom Java software. This view will enable the puppeteer to interact with the audience by supplying a live feed of the viewers along with a reticle for selecting which audience member the digital puppet should attend to. Similarly, the puppeteer will be given a first-person view of the virtual environment to attend to artifacts that the audience should focus on. An overhead map view will allow the puppeteer to select where he/she wishes the puppet to travel throughout the virtual environment. Once a location is selected, the puppet will begin moving towards that area. This simplifies the puppeteer's task by removing the need to focus on the directional movement of the puppet. Additionally, another screen may eventually be added in order to provide a third-person view of the virtual environment. This view would be what the audience sees, provided to allow the puppeteer to see what his/her actions are producing. The puppeteer will be given Wii® remotes (Wiimotes®) or a mouse in order to interact with these views.

Procedure

Fitts' Law was used to form the basis for determining which controller (the Wiimote® or the mouse) and which screen switching method (toggling or pointing) is the most intuitive and effective form of control for the user of the interface (Fitts, 1954). Custom Java software was developed to test how quickly participants can click from one target displayed on the screen to the next, using four different forms of control. These forms of control were:

1. Mouse Toggle - the mouse is used as the input device for target selection, peripheral screens are brought into focus using the "left," "down," and "right" arrow keys on the computer's keyboard
2. Mouse Point - the mouse is used as the input device for target selection, peripheral screens are brought into focus by pointing at a colored band at the bottom edge of the focal display which corresponds to the correct peripheral display
3. Wiimote® Toggle - the Wiimote® is used as the input device for target selection, peripheral screens are brought into focus using the "left," "down," and "right" arrow keys on the Wiimote® d-pad
4. Wiimote® Point - the Wiimote® is used as the input device for target selection, peripheral screens are brought into focus by pointing at a colored band at the bottom edge of the focal display which corresponds to the correct peripheral display

The targets, which were displayed as red dots, randomly changed size and location on the screen each time one was clicked. The distance traveled to the next target and the time it took for a user to focus and click on the target were calculated and used to determine the index of difficulty (Fitts, 1954). "Targets" were red circles with radii ranging from fifteen to thirty-five pixels. Custom Java software randomly generated the screen on which a target was displayed. In the same way, the location of the target within the screen was also randomly generated. The larger, main focal display was established with a projector, shown just above the peripheral displays. An image of the setup is presented in Figure 1.

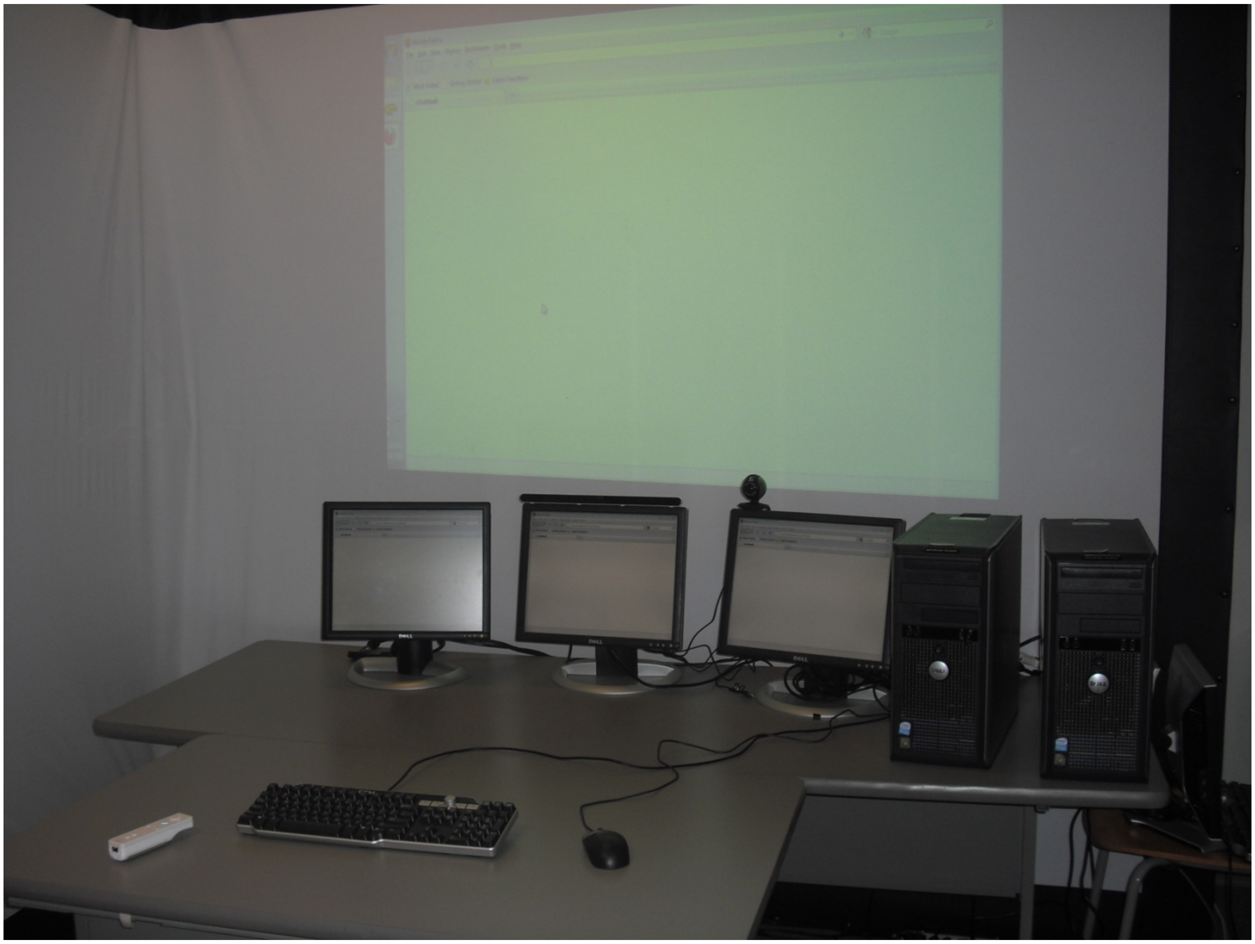
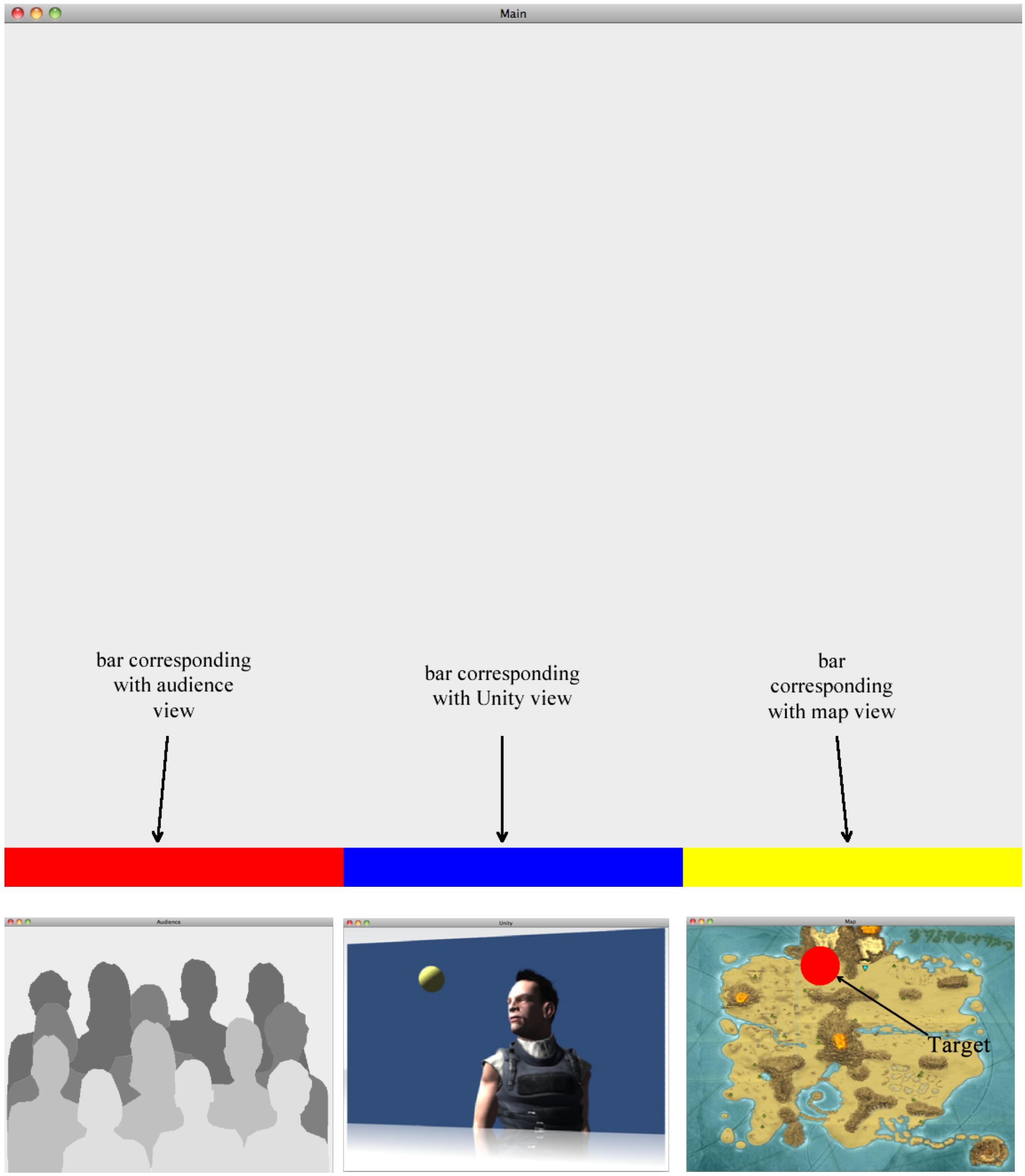


Figure 1. This figure illustrates the design of the system.

The trials were run on a PC with a Windows 7 operating system. The three peripheral displays were shown on three separate 15" monitors by using a device called Matrox TripleHead2Go. This device allows for the video source to be split across three different screens. The connection between the Wiimote® and the computer was established via bluetooth. The native Windows bluetooth drivers proved to be ineffective, and so the bluetooth driver was provided by an application called Blue Soleil. An application called GlovePIE (Glove Programmable Input Emulator), created by Carl Kerner, was used to allow the Wiimote® to control the mouse pointer. This open source project was originally intended to emulate joystick and keyboard input with virtual reality gloves, but it has since been expanded to emulate multiple other devices. A custom script was created and run through the GlovePIE application which allowed the Wiimote® to control the mouse pointer through infrared motion tracking (LEDs were created by a wireless Wii® sensor bar).

Users were asked to perform three trials with each form of control, fifty clicks in each trial, giving a total of one hundred and fifty data points per participant per technique. This was done so that the data set would be large enough reduce possible skewed results. Prior to actual tests, participants were given "training" sessions where they were given the freedom to click on any number of targets until they felt comfortable with the control scheme. The purpose of these training runs was to reduce the possibility that any of the data points would be collected prior to a participant being properly oriented with any one form of control.

Figure 2 imitates what a participant might see when the application first begins:



Audience View　　Unity Environment View　　Map View

Figure 2. This figure shows what a user would see at system startup.

After each form of control had been tested, participants were given a post session survey in which they were asked a series of questions, as depicted in the images below (Note that

"Warp" here means "Toggle." This was explained to the participants prior to administering the exit survey).

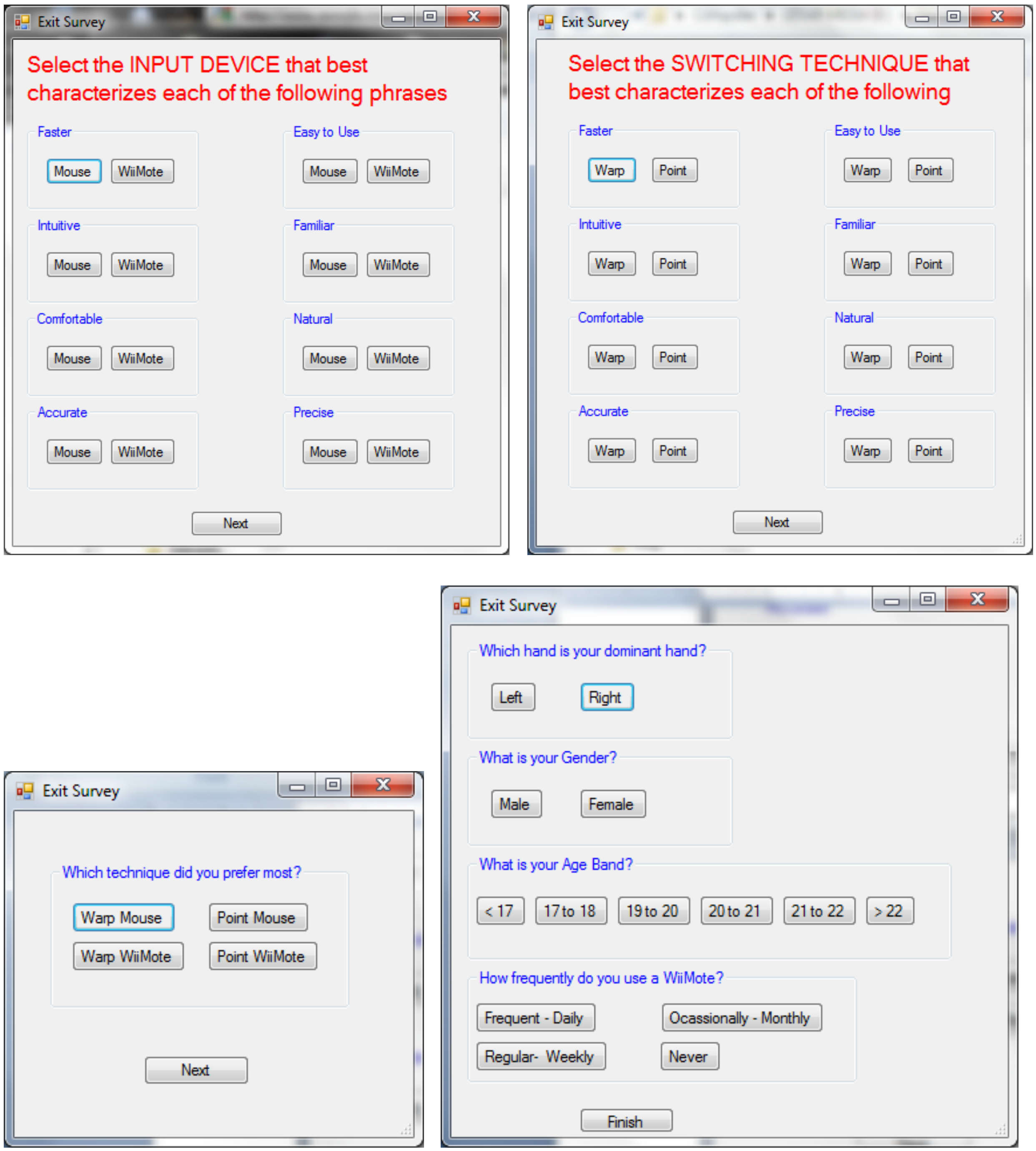


Figure 3. This figure shows the questions given to participants in the exit survey.

Participants

Data from ten participants (each a student currently enrolled at the University of Northern Iowa) was collected for the purpose of this study, for a total of 1500 data points per technique. The study was originally designed to collect data from twelve participants. The order in which the different forms of control were tested was not the same for every subject. There were four different orders in which a subject was to be tested, one for every three participants.

Table 1

Technique Order Per Participant

Key:
Mt = Mouse Toggle
Mp = Mouse Point
Wt = Wiimote® Toggle
Wp = Wiimote® Point

Subject #	Order
1	Mt Mp Wt Wp
2	Mt Mp Wt Wp
3	Mt Mp Wt Wp
4	Mp Mt Wp Wt
5	Mp Mt Wp Wt
6	Mp Mt Wp Wt
7	Wt Wp Mt Mp
8	Wt Wp Mt Mp
9	Wt Wp Mt Mp
10	Wp Wt Mp Mt
11	Wp Wt Mp Mt
12	Wp Wt Mp Mt

However, by the end of the study only ten of the subjects' data could be used. One participant failed to show up to run the trials, and software anomalies experienced during one of the trials rendered the data points collected from that participant useless. Thus, only ten orderings displayed above were implemented. Subject 9 experienced the anomalies, so that order was not used, and the twelfth ordering was also unused. Participants were given $10 gift cards for their time.

Analysis

Each time a user clicked on a target, data were collected and later written to a file. This collected data contained: the size and location of the target, a timestamp for when the target was clicked, the point where the user clicked on the target, which peripheral screen the user selected, and a timestamp for when this screen was selected. If the form of control was a pointing method, then the point where the user entered the colored bar was also collected for data analysis. The screen names and times selected for any incorrect screens chosen prior to the one in which the target resided were also recorded. Data were recorded in comma-separated value (CSV) file.

This CSV file was then imported into a Microsoft Excel spreadsheet for further evaluation. Each form of control for all ten of the participants was evaluated separately. For toggling forms of control, the points that were evaluated included: the total distance traveled from one target to the next, the total time it took the participant to select the next target, the time it took the user to make a transition (how long it took them to realize which screen held the target), and the time it took the participant to select the target once he/she had selected the correct screen. Pointing forms of control were evaluated slightly differently because they had to account for the user pointing first to the appropriate bar, and then coming up from that point to select the target. For these forms of control, the point where the user touched the bar was recorded in the CSV file. This point was then evaluated in excel in order to obtain the total distance the subject had to travel from one target to the next. For each form of control, the average time it took the participant to select the correct target was determined for each of the three trials. The results were then plotted on a regression plot. The results for each subject are shown in figures 4-13.

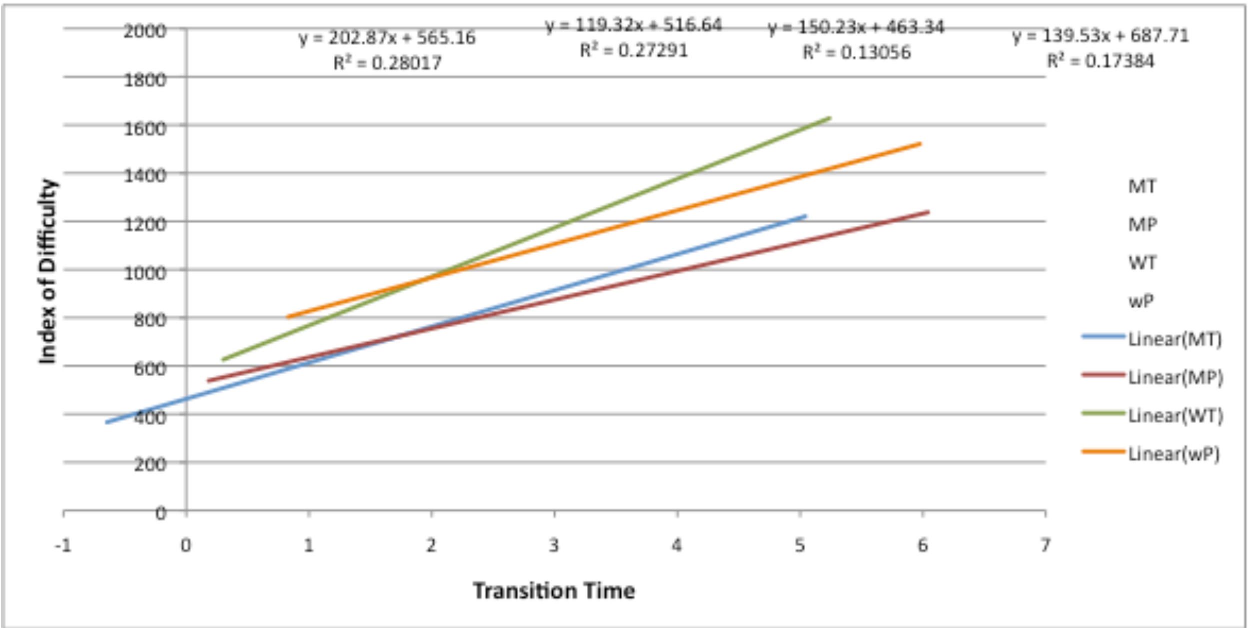


Figure 4. Subject 1

The x-axis represents the index of difficulty, as defined by P. M. Fitts: ID = $\log_2(2A/W)$, where *A* is the distance from the previous click to the center of the next target, and *W* is the width of the target (the radius of the circle in this experiment) (MacKenzie, 1992). The y-axis represents the transition time. Note that there is a negative index of difficulty with this subject's data (and with some of the following subjects as well). This seemingly impossible occurrence can take place when a target appears at a point that is closer to the last click than the target circle's radius. This means that a user could select a target without moving the cursor. Also note that the r-squared values for these individual analyses are rather low. This is because the data are not “bucketed,” meaning they are not condensed. For example, if the data were bucketed, any data points with a given index of difficulty between 1.2 and 1.4 would provide a single data point.

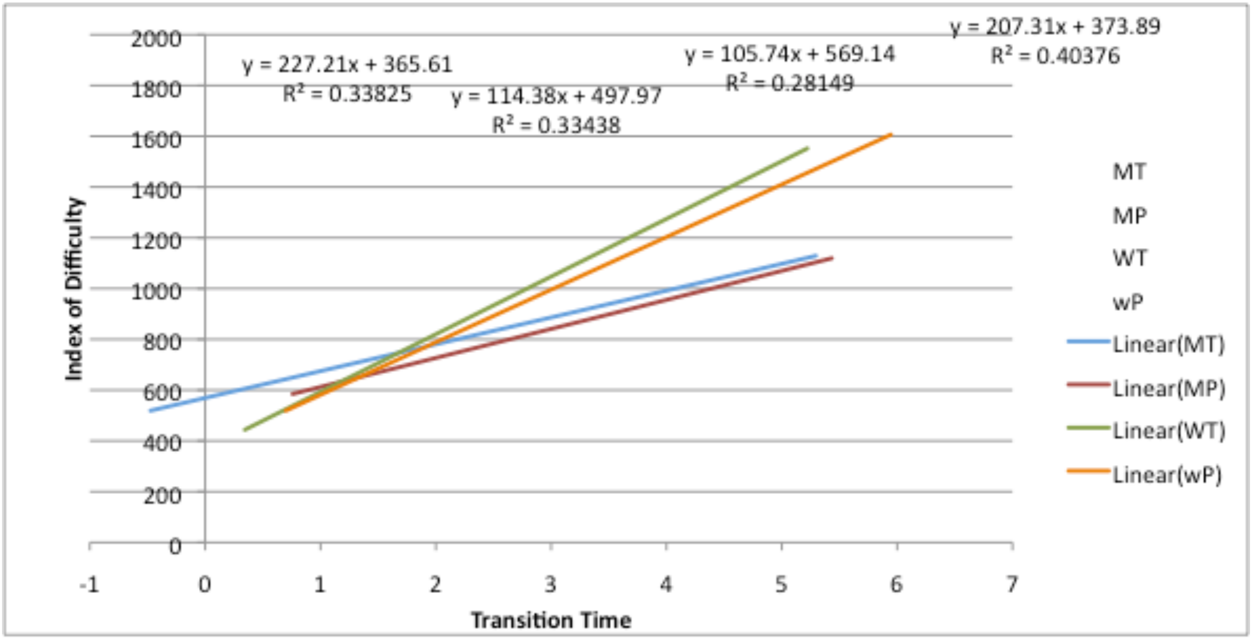


Figure 5. Subject 2

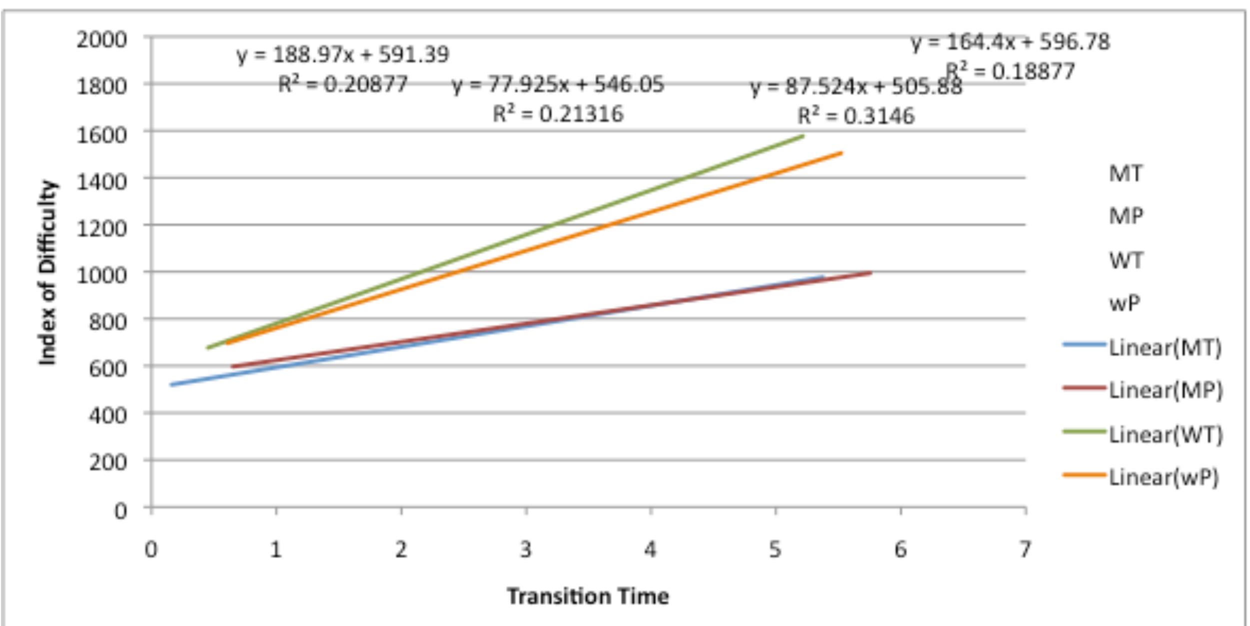


Figure 6. Subject 3

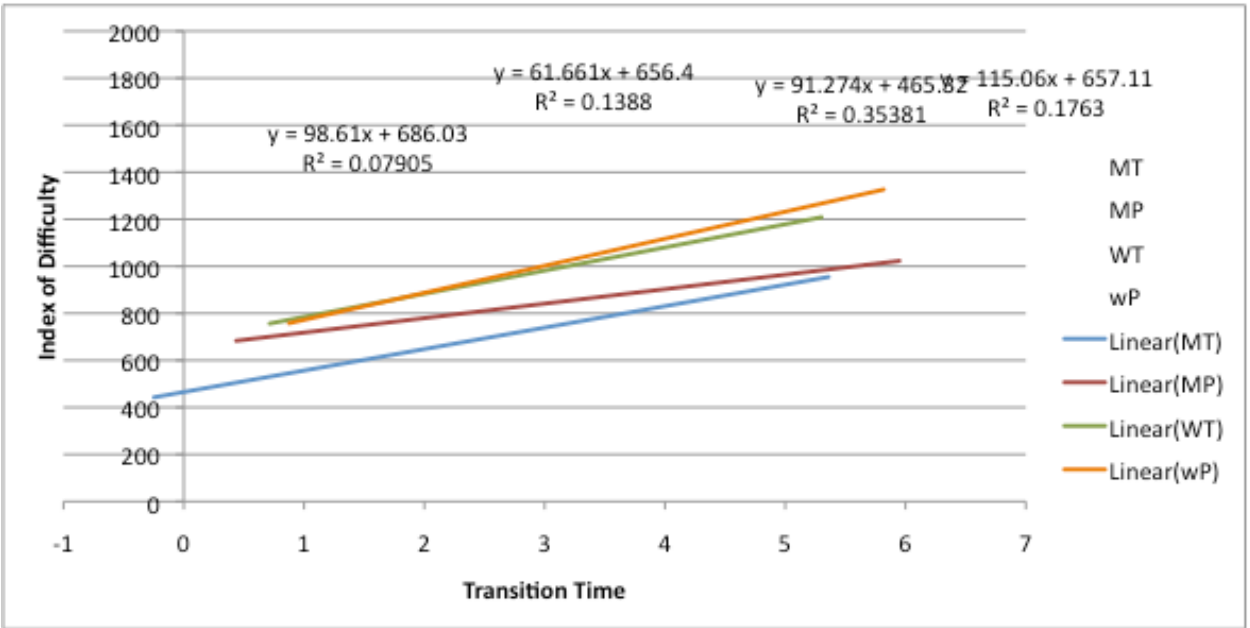


Figure 7. Subject 4

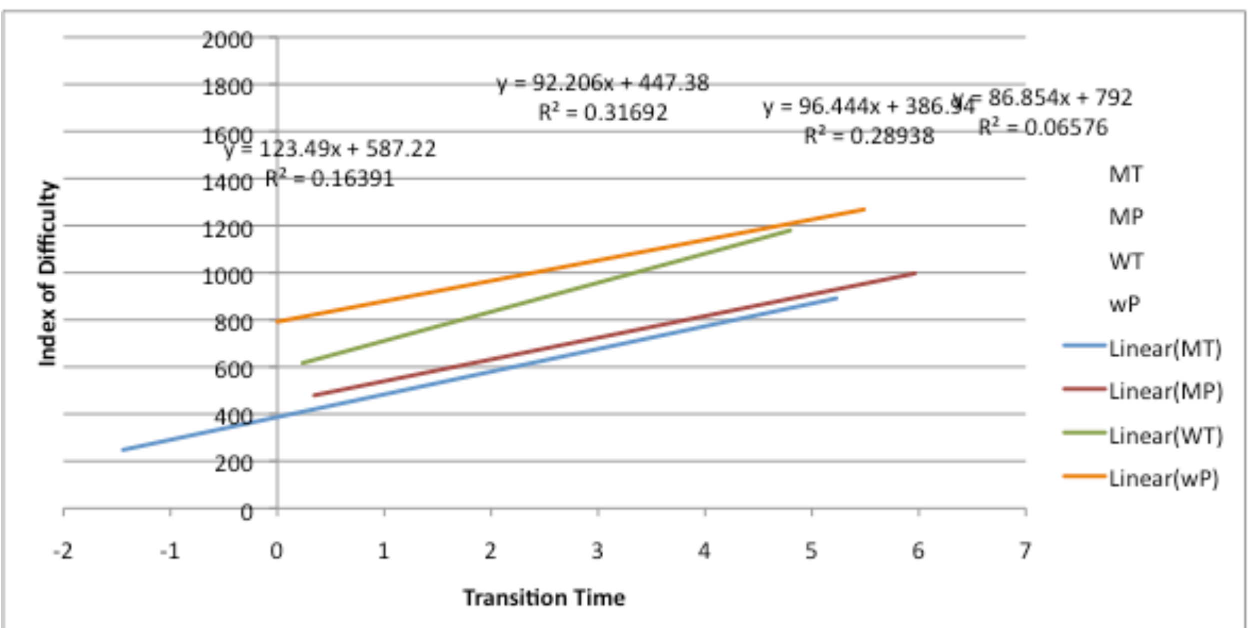


Figure 8. Subject 5

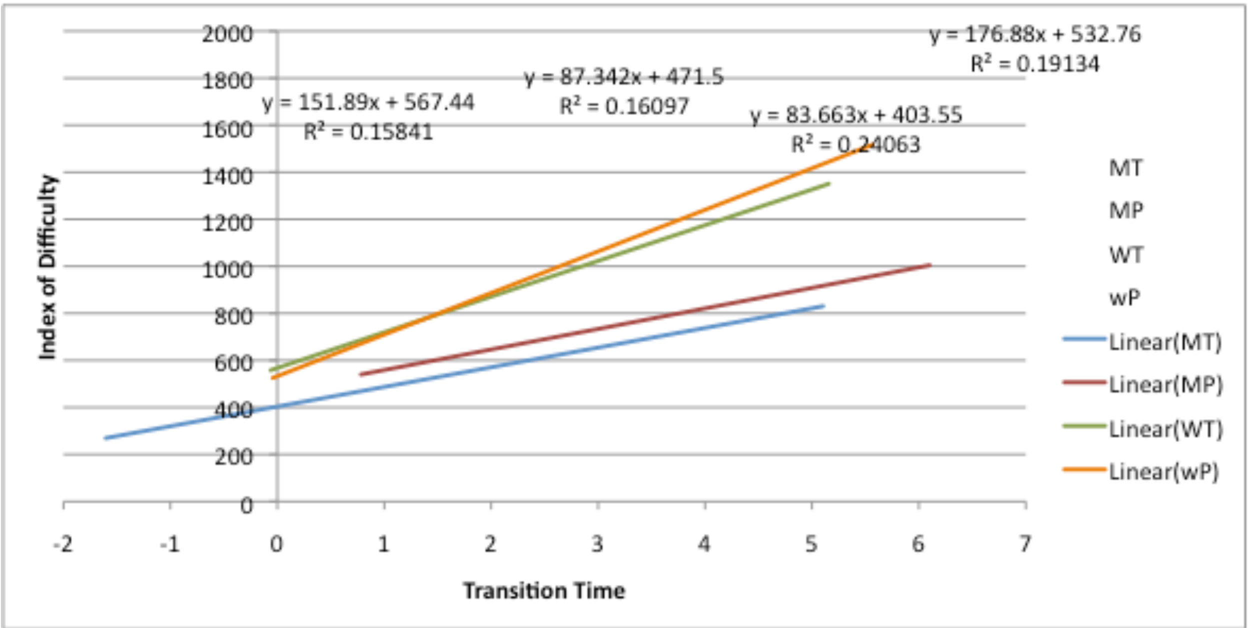


Figure 9. Subject 6

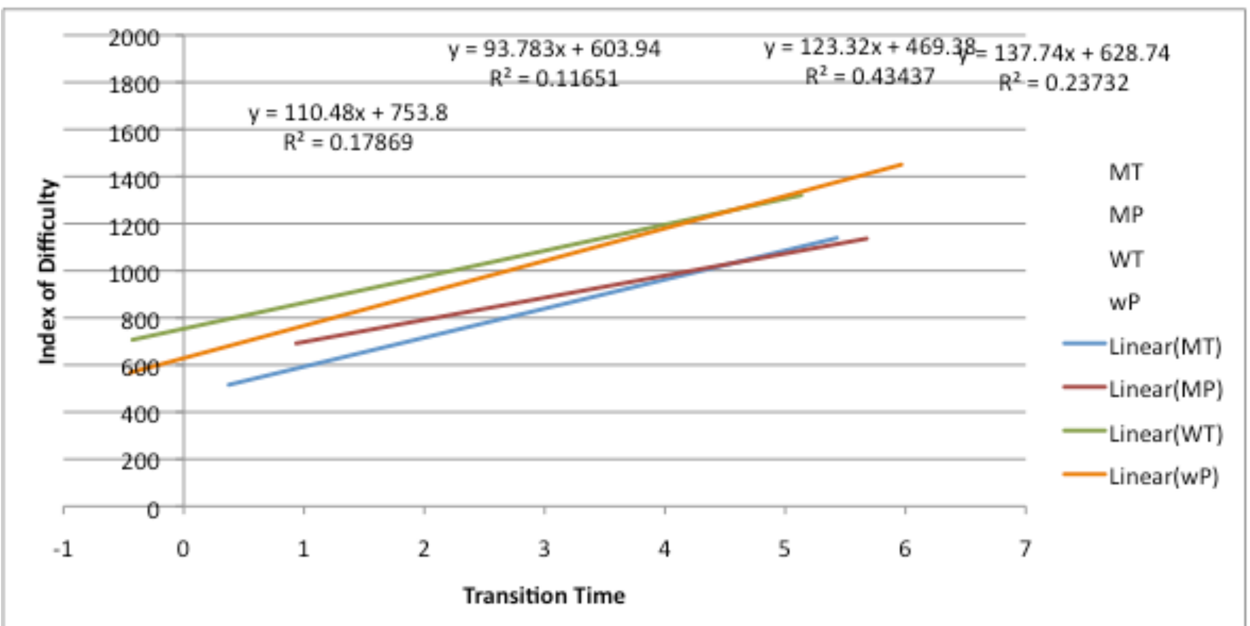


Figure 10. Subject 7

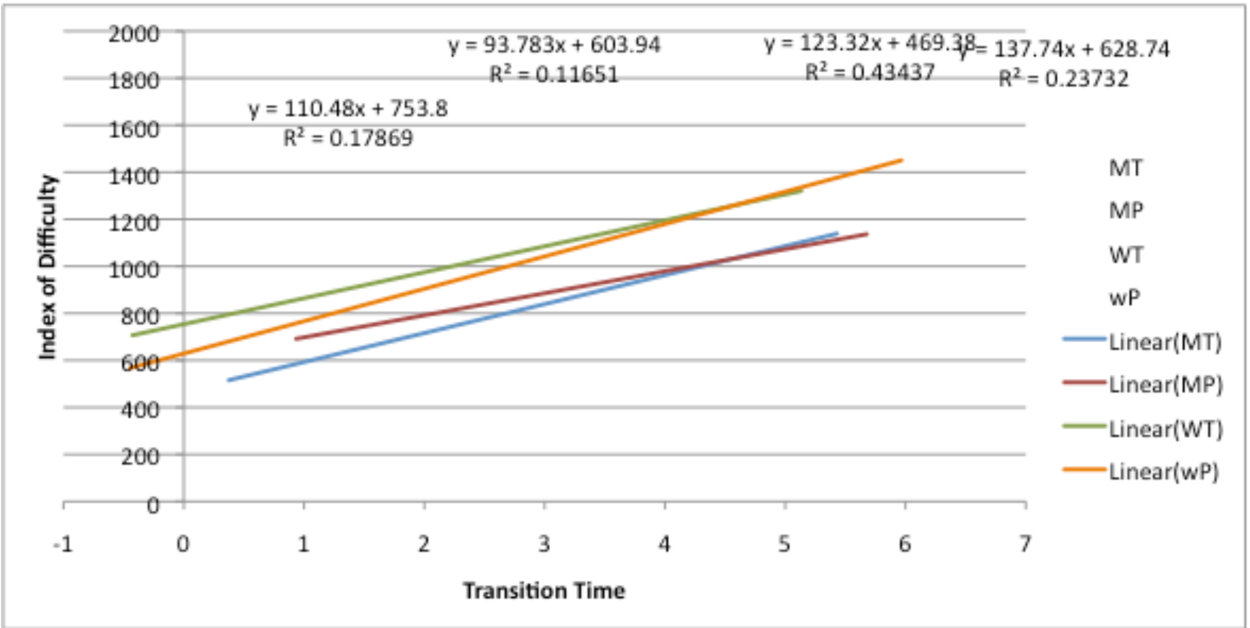


Figure 11. Subject 8

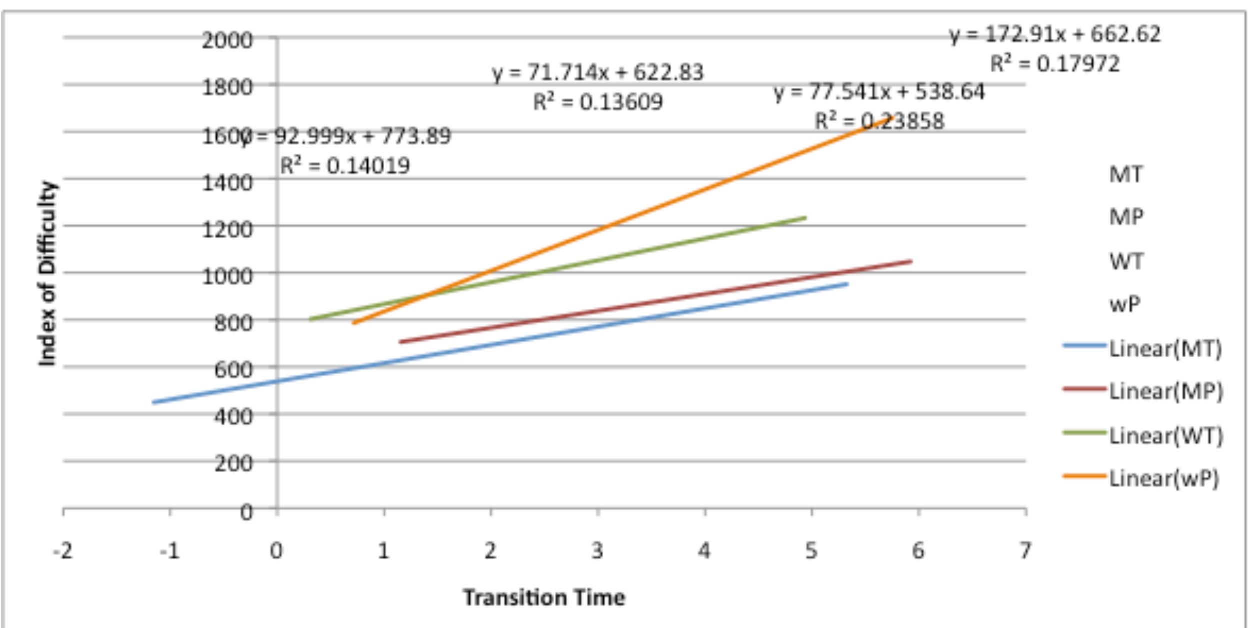


Figure 12. Subject 10 (Subject 9's data was discarded)

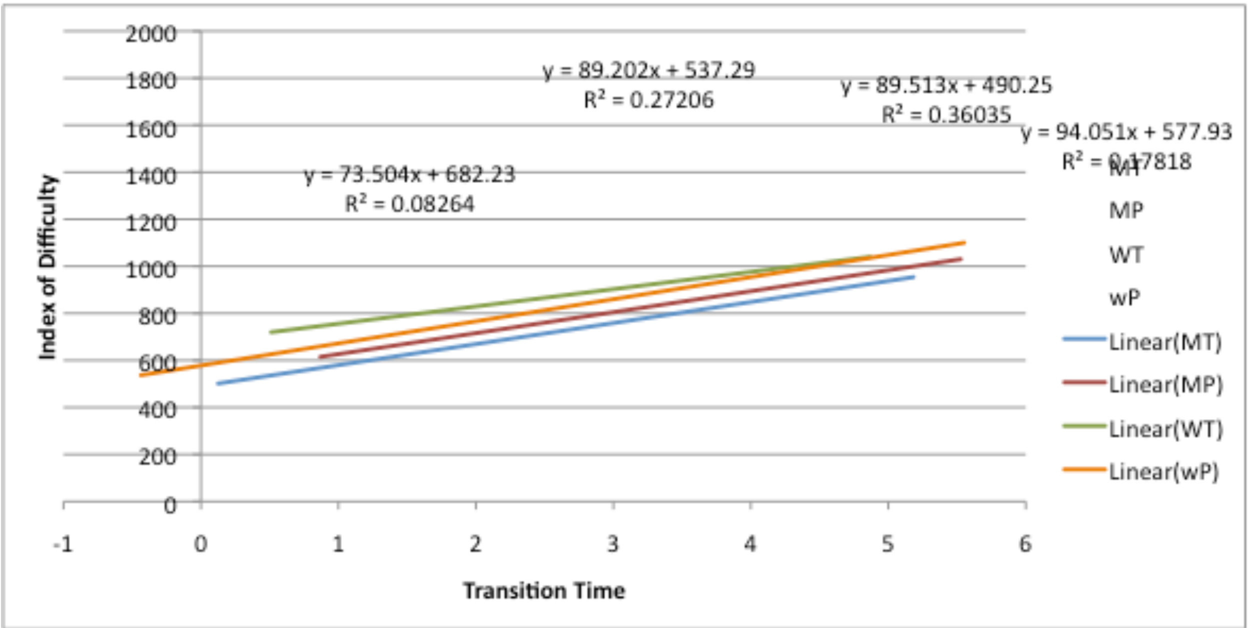


Figure 13. Subject 11

After the results from each participant were collected, an aggregate analysis of the data was created, as shown in Figure 14.

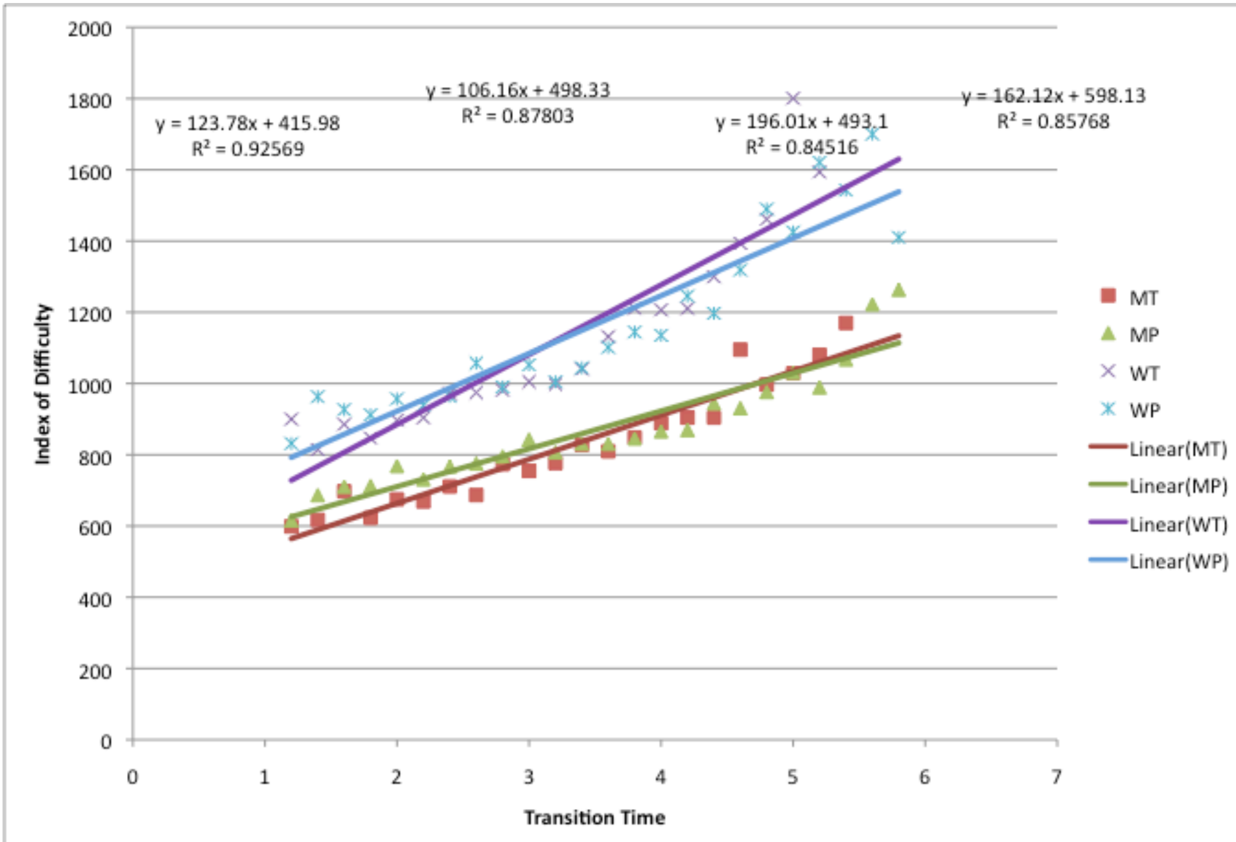


Figure 14. An aggregate analysis of the subject data

Again, the x-axis represents transition time and the y-axis represents the index of difficulty. The data points shown in Figure 14 are bucketed groups of data points, combined by basis of the index of difficulty. Data points with an index of difficulty less than 1.2 were excluded in the aggregate analysis. The chart shows a sizeable difference for the transition times between the mouse forms of control and the Wiimote® forms of control. The transition times for mouse methods were noticeably lower than those of the Wiimote® methods.

There are a few things to note in this analysis:

a) The r-squared values are really high. This shows that Fitts' Law is holding; there was a (strong) linear relationship between the ratio of distance over size of the target and the selection time. This correlation was highly significant, and thus meaningful for the design of the interface.

b) When comparing the lines, the first thing that to observe is that there is a gap between the red/green lines and the blue/purple lines. This gap relates to a speed difference between the Wiimote® based techniques and the mouse based techniques. The mouse-based techniques were generally faster, and this difference is statistically significant. Moreover, it is not a difference without distinction; one can see that this gap is on the order of a half second. This may not initially sound like a very large difference, but if every click took a half second longer, the performance difference would be significant.

c) The slopes of the Wiimote® lines are steeper than the slopes of the mouse lines. This could mean that it is more difficult for the Wiimote® to acquire targets that are smaller, or it could mean that it is more difficult for the Wiimote® to acquire targets that are further apart. Further study is needed in order to understand which of these is true (although from watching the study it would appear to be the smaller ones). In either case, this steeper slope is another indication that the device is harder to use.

d) A final observation is that the lines for both devices cross depending on the transition type. This interaction was not found to be significant, but it might merit further study. (If this were the case, it would mean that the transition type was impacting the user's ability to acquire "difficult" targets.)

In addition to the regression plots shown in Figures 4-14, paired two sample t-Tests were used to compare the effectiveness of each form of control relative to the other methods. They were utilized to evaluate how long it took for a user to transition to the correct screen in which the target was located. In general, tests of this form are often used to determine whether two variables are related. If they are related, this test can also reflect how a change in one variable might affect the other. For example, if variable one is increased, how does this affect variable two?

The results of the two sample t-Tests show that the mean transition time of the Wiimote® pointing method was approximately 138 milliseconds greater than that of the mouse toggling method, with a p-value of 0.005. The mean transition time for the Wiimote® toggling method was also greater than that of the mouse toggling method by approximately 77 milliseconds, with a p-value of 0.017. The mouse pointing method was slower than the mouse toggling method as well, by approximately 69 milliseconds, with a p-value of 0.001. The final statistically significant comparison to note from these results is that the Wiimote® pointing method produced higher mean transition times than the Wiimote® toggling method, by approximately 60 milliseconds, with a p-value of 0.04.

These results from the two sample t-Tests illustrate that, within devices, toggling was always faster than pointing. They also show that the mouse toggle form of control performed significantly better than any other form of control (the transition times were quicker), and the Wiimote® toggle method performed better than the Wiimote® point method (assuming a significance level of five percent).

Post Session Survey Results and Demographics

Of the ten participants who successfully completed the trials, six were female and four were male. Three of them were left-handed, and the other seven were right-hand dominant. When asked how frequently they played the Wii® (used to determine the level of familiarity with a Wiimote®), two subjects responded with "Never," seven responded with "Occasionally," and one responded with "Regularly." When asked which method was preferred, the most frequent response from the participants was the mouse toggle form of control. Although the majority of participants (eight of the ten) favored this method, one participant chose Wiimote® toggle form of control and another chose the mouse point form of control. Even though the data collected from the post session survey is not enough to be statistically significant, it still remains relevant to the study and was kept in consideration.

Conclusion

Discussion

Interestingly, the collected data contradicts the aforementioned hypotheses. Prior to running the trials, it seemed intuitive that the most effective order of control would be the Wiimote® pointing method. Researchers such as Cheng and Pulo (2003) claim that direct control devices (such as the Wiimote®) are more effective when interacting with large-screen displays. Theoretically, direct control devices should decrease the cognitive load of the user and allow for increased reaction times (Cheng & Pulo, 2003). The results of this research, however, show that this is clearly not always the case.

One of many reasons as to why this is so might be due to the familiarity of people in today's society with a mouse. The typical human being interacts with his/her computer on a daily basis, often for hours a day. The mouse is an extremely familiar tool that most people are comfortable with using. The Wiimote®, on the other hand, is far less familiar. Indeed, only one participant claimed regular Wiimote® usage, and two others asserted they never used the device at all. Although reasons such as this can possibly be used to help explain the results of this experiment, they cannot adequately justify the findings. In fact, researchers MacKenzie and Jusoh (2001) performed a similar experiment in which pointing devices were compared with the mouse. They did not believe that a user's expertise with the mouse was sufficient to explain increased performance (MacKenzie, 2001). There may be more compelling reasons at work here.

One of these reasons is given by the work of K. M. Baird, E.R. Hoffman, and C. G. Drury (2002). The focus of their study was to determine the effect of probe length on Fitts' Law analyses such as this one. What they found was that probe length directly influences MT (transition time). To be more precise, the effect of the probe length appears to be a coefficient of the index of difficulty. In other words: the longer the probe, the greater the transition time will be. When comparing the results of their study to the results of this one, their explanation appears to fit. The "probe length" for the Wiimote® in this experiment was approximately eight feet. The "probe length" for the mouse was essentially zero. Since the probe length for the Wiimote® is far greater than that of the mouse, Baird et al. (2002) would claim that the average transition time should indeed be larger. They even go so far as to state: "...combinations of long probes and accurate target aiming should be avoided" (p. 6). Their findings can help explain and support the results of this research.

Based on the results of this study, the most effective interface for the proposed virtual puppetry system would use the mouse toggling form of control. It should be noted, however, that these results might not apply to every similar system. In addition, while the faster transition times found with the mouse input methods suggest a greater level of situation awareness, they do not necessarily guarantee it. As noted by Endsley (1995), these two variables are linked, but correlation does not necessarily mean causation. Situation awareness depends on a number of factors that would be difficult to include in a single study. Even so, the results of this experiment shall be a useful tool in the design of the virtual puppet interface. If nothing else, they show that what may seem to be the most intuitive form of control is not always so. These data give a respectable base for further work in this area to refine upon. They provide a justification for deciding to use the mouse toggle form of control in the virtual puppet interface, and are an important contribution to work in virtual heritage sites.

Limitations

Of course, this study had its limitations. For instance, there were limited demographics on the population sample since participants had to be students of the University of Northern Iowa. It is difficult to determine if this sample would be truly representative of the entire population. This study is also not entirely representative of what the future system will be like. The images given to users were still frames, whereas during an actual implementation of the puppet control interface the operator would be dealing with moving scenes. The target, being a bright red dot, may have been more conspicuous than what the puppeteer would experience as well. Even though these limitations do exist, they do not appear strong enough to discredit the findings of this research.

Future Work

It would be beneficial for future studies on this subject to work with larger population and sample sizes so that they are not limited in the same way as this experiment. The forms of control could also be retested on the completed interface, which would give more accurate results than the prototype constructed for this research. More control devices, such as a joystick or analog controller could be tested against the mouse and Wiimote®. One such control device, the PlayStation® Move controller, has the potential to outdo the Wiimote® in precision and accuracy with its three-dimensional motion tracking (the Wiimote® is limited to two dimensions); however, there are currently no methods for using this device in the same manner as a mouse. Once these are developed, it would be interesting to redo this study with the PlayStation® Move controller.

References

Baird, K. M., Hoffman, E. R., & Drury, C. G. (2002). The effects of probe length on Fitts' law. *Applied Ergonomics, 33(1),* 9-14.

Campbell, B. A., O'Brien, K. R., Byrne, M. D., & Bachman, B. J. (2008). Fitts' law predictions with an alternative pointing device (Wiimote®). *Human Factors and Ergonomics Society Annual Meeting Proceedings, 52*(19), 1321-1325.

Cheng, K., Pulo, K. (2003). Direct interaction with large-scale display systems using infrared laser tracking devices. *Australasian Symposium on Information Visualisation, Conferences in Research and Practice in Information Technology, 24*.

Czerwinski, M., Smith, G., Regan, T., Meyers, B., Robertson, G., & Starkweather, G. (2003). Toward characterizing the productivity benefits of very large displays. *Interact Conference,* 9-16.

Dede, C. (2009). Immersive interfaces for engagement and learning. *Science 2, 323*(5910), 66-69.

Devine, K. (2007). Place and time: creating contextualised presence for virtual heritage. *Proc. 13th Intl Conference on Virtual Systems and Multimedia,* 65-72.

Endsley, M. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 37*(1), 32-64.

Fitts, P. M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology 47*(6), 381–391.

Forlines, C., & Balakrishnan R. (2008). Evaluating tactile feedback and direct vs. indirect stylus input in pointing and crossing selection tasks. *CHI 2008 Proceedings · Tactile and Haptic User Interfaces,* 1563-1572.

Jacobson, J. Ancient architecture in virtual reality: Does immersion really aid learning? *PhD Thesis*, University of Pittsburgh, 2008.

MacKenzie, I. S., Buxton, W. (1992). Extending Fitts' law to two-dimensional tasks. *CHI '92 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems,* 219-226.

MacKenzie, I. S., Jusoh, S. (2001). An evaluation of two input devices for remote pointing. *EHCI '01 Proceedings of the 8th IFIP International Conference on Engineering for Human-Computer Interaction,* 235-250.

Mateas, M. (2001). Expressive AI. *Leonardo: Journal of the International Society for Arts, 34*(2), 147-153.

Ulicny, B., & Thalmann, D. (2002). Crowd simulation for virtual heritage. *Proc. First International Workshop on 3D Virtual,* 28-32.