The Use of the Kinect in a Medical Serious Game

Matthew Hackett	Howard Mall	Mark Mazzeo
ARL-HRED STTC	ECS	ARL-HRED STTC
Orlando, FL	Orlando, FL	Orlando, FL
Matthew.G.Hackett@us.army.mil	howardmall@ecsorl.com	Mark.V.Mazzeo@us.army.mil

ABSTRACT

Interaction within a virtual environment is a key factor relating to a positive user experience. The current industry standard for interaction and character control is via mouse and keyboard or a game controller. Unfortunately, neither of these input mechanisms represents a natural input modality for a trainee. In the training domain, this disconnect can potentially result in negative training transfers and user frustration. Fortunately, recent advances in gesture and motion based control have opened up new avenues for intuitive and natural control schemas. Through a current research effort, the Tactical Combat Casualty Care Simulation (TC3Sim) serious game has been updated to use the Kinect motion sensor as an input device. The integration of the Kinect interface with a medical training game has revealed significant findings regarding its usability and performance. The research has also found general guidelines for the use of a natural user interface regarding poses, body positions, dynamic gestures, and menu interactions. This paper details these guidelines as well as the benefits and drawbacks of a natural user interface for a serious medical game. The changes made to the user interface to overcome many of those drawbacks are presented, including simplification of the control schema and certain game play alterations. Results of an initial usability study are also presented, focusing upon user interaction with the game and satisfaction metrics to identify the feasibility of the Kinect sensor in other medical simulation efforts.

ABOUT THE AUTHORS

Matthew Hackett is a science and technology manager for the Medical Simulation Research Branch of the U.S. Army Research Laboratory, Human Research and Engineering Directorate, Simulation and Training Technology Center (ARL-HRED STTC). He manages a variety of projects including the medical holography research and virtual patient research efforts. As a science and technology manager, he oversees these research efforts and conducts test and evaluation to determine their efficacy in the simulation and training domain. Prior to his work with ARL, Mr. Hackett trained to be a government engineer while working as an engineering intern at the Program Executive Office for Simulation, Training and Instrumentation and worked within PM Training Devices. Mr. Hackett received his Bachelor of Science in Computer Engineering from the University of Central Florida and his Masters of Science in Biomedical Engineering from the University of Florida. Mr. Hackett subsequently received his Masters of Science in Modeling and Simulation and is currently pursuing his Ph.D. in Modeling and Simulation at the University of Central Florida.

Howard Mall is Vice President of Engineering at Engineering and Computer Simulations, Inc. He has spent the last nine years at ECS building various kinds of training systems. He has lead efforts for the Navy to develop training solutions deployed on cell phones and hand-held computers. For the Army, he delivered the Tactical Combat Casualty Care (TC3) Simulation used by combat medics to learn triage and medical decision-making on a virtual battlefield. He has led several virtual world and game-based simulation projects for a myriad of federal agencies and commercial concerns. He currently oversees multiple engineering efforts at ECS both mainstream and on the fringe of the state-of-the-art.

Mark Mazzeo is an Engineering Technician for Medical Simulation Technologies at the U.S. Army Research Laboratory, Human Research and Engineering Directorate, Simulation and Training Technology Center (ARL-HRED STTC). As an intern, he takes an active role in the management of current efforts by accompanying Science and Technology Managers out in the field, assisting with documentation, data collection, and experimental design. He prepares and reviews contractual documents, performs technology demonstrations and maintains laboratory equipment, and explores new areas of research within Medical Simulation through literature reviews. Mr. Mazzeo is currently pursuing a B.S. in Industrial Engineering from the University of Central Florida, and plans to obtain a M.S. in either Systems Engineering or Modeling and Simulation.

The Use of the Kinect in a Medical Serious Game

Matthew Hackett
ARL-HRED STTC
Orlando, FL
Matthew.G.Hackett@us.army.mil

Howard Mall
ECS
Orlando, FL
howardmall@ecsorl.com

Mark Mazzeo
ARL-HRED STTC
Orlando, FL
Mark.V.Mazzeo@us.army.mil

INTRODUCTION

Virtual medical simulations present a variety of challenges: complex physiology models, branching decision trees, a massive problem space, and a realistic simulation interface. The interface requires the capability of accessing a wide variety of medical actions while minimizing user confusion. Furthermore, virtual medical simulations strive to emulate the interaction between a medical practitioner and patient. Unfortunately, the control schemas involving a mouse and keyboard or game controller do not achieve this goal. A simple click or button press is unable to simulate the effect of reaching out and touching a patient or applying a bandage. The problem compounds when trainees focus more upon learning an interface than the skills intended. This disconnect between the actions performed and the interface options may be more than just a distraction and result in negative training transfers. The level of immersion presented by a simulation greatly impacts the user (Mania & Chalmers, 2001). Therefore, the goal of any virtual medical simulation is a natural, intuitive interface that immerses a trainee into the virtual environment.

Military training requirements are extensive for a medic or combat life saver, to include balancing combat readiness courses with medical training. While their schedule is crowded with events and courses, in many instances there is significant downtime. To maximize training efficiency a reduction of downtime is necessary. In addition to improved efficiency, trainees are more satisfied during training events with limited downtime (Solstad, 2011). Development of technology to fill the downtime and reinforce training objectives is a central research goal within the Army. The development of the TC3Sim platform targeted downtime in this fashion. The TC3Sim is a 'first person thinker' placing the user in the eyes of a Combat Medic. The game forces trainees to properly triage casualties and treat them according to the principles of Tactical Combat Casualty Care (TC3) (Butler, 2010). Studies showed that the TC3Sim platform was effective at improving learning outcomes when used in conjunction with traditional classroom instruction (Sotomayor, 2010). The TC3Sim platform improved significantly over the years. The graphics of the game improved substantially switching from Gamebryo to the open-source Unity graphics engine. A high fidelity physiology model can be attached to increase the fidelity of patient responses to treatments (Sotomayor, 2012). A great deal of research and development improved the platform; however, the interface still centers on the mouse and keyboard, prompting research into a natural interface to improve immersion. The TC3Sim platform is an excellent fit for research on such an interface, due to its focused scope of medical treatments and lack of complex interface operations, such as drop-down menu navigations. After considering numerous interface options, the Microsoft Kinect was chosen as the interface hardware due to its robust Software Development Kit (SDK), excellent gesture recognition capability, and relatively low cost. The path to incorporating the Kinect within TC3Sim required technical research into the Kinect, software development creating a gesture library, and integration of the Kinect software with the interface scheme of TC3Sim. After completion of these goals, we conducted an initial usability study to determine user satisfaction metrics and to identify current shortcomings in the interface design. The intended result of a successful natural user interface is high levels of user satisfactions, high levels of immersion, and ultimately improvements in training.

TECHNOLOGY BACKGROUND

A Natural User Interface (NUI) describes a class of devices that allow humans to interact with computers in ways that are easy to remember and execute. The form of interaction has typically been ascribed to things that humans already naturally do like speech, touch, and gesture. The Kinect is a NUI that the Microsoft Corporation introduced to the market in November of 2010, intended as an accessory for the XBox 360 gaming platform and was their answer to Nintendo's Wii controller that used a motion-sensing remote control to add gestures to games. It was promptly hacked when an individual on the internet released an open source USB driver for the Windows operating system six days later (Lowensohn, 2011). It quickly became a boon for human-computer interaction and robotics researchers. A low-cost, highly capable optical depth sensing system was now available for all kinds of people to use for research or play.

The Kinect is composed of a camera, a depth sensor that uses an infrared laser projector, and an array of microphones. These sensors combine with software drivers to support whole-body motion capture, voice recognition, and facial recognition. The ability of the Kinect to recognize gestures of an individual's body in its frame of view is of prime importance to the work described in this paper.

Along with the Kinect hardware, Microsoft also released several game titles using the new interface. These games use your whole body to navigate and control the action on the screen. The promise of the Kinect became evident in games where the system accurately measured player dance motions and compared them against on-screen dancers. The best games for the Kinect are usually simple and designed from the ground up to incorporate large gestures as a primary mechanic.

Microsoft later released official drivers and new Kinect hardware specifically to work with their Windows 7 operating system in June of 2011. It includes a significant number of programming examples and the cameras have been altered so the Kinect can be used with the operator closer to the device. The TC3Sim Kinect development used the latest software and drivers.

Kinect Integration Technical Effort

The Kinect libraries and control capabilities were integrated into the TC3Sim base software code. The Kinect toolkit contains a set of C++ examples and application programmer interfaces that allows gestures to be easily implemented as user input. The incorporation of the Kinect hardware was not as significant a task as redesigning the TC3Sim user interface to be more "natural".

The "normal" TC3Sim interface works quickly and intuitively with a keyboard and mouse. The mouse controls the player's point of view. Move the mouse up and down to look up and down; move the mouse left and right to look left and right. Keyboard keys translate the player through the environment. A reticule appears in the center of the screen. The reticule aims and fires the weapon but also serves as a contextual menu interface. The user points the reticule on an object, person, or body area in the virtual environment, right clicks, and then selects an option from choices arrayed in a circle around the screen. The menu allows the user to select all the different medical interventions or tests on synthetic casualties, to secure a weapon, or to request help from a non-player character. This is quick and intuitive without the need to create a system with multiple modes.

The original Kinect implementation of the TC3Sim attempted to build upon this well established user interface paradigm. However, the research found that an interface optimized for ease-of-use with a mouse and keyboard proved to be much less intuitive using the Kinect-style of NUI. In the initial Kinect version of TC3Sim, the user navigates through the environment by moving their body. The user rotates through shoulders left and right at the waist to rotate the first person point of view. To look up and down, the user leans forward and backward at the waist. To move forward, the user steps forward. To move back, the user steps back. As with the original TC3Sim, the user positions the reticule over the intended object or body part. To access the menu, the user lifts a hand. The position of the raised hand controls the position of the cursor. To "click" on a menu option, the user raises and lowers their other hand (Figure 1).

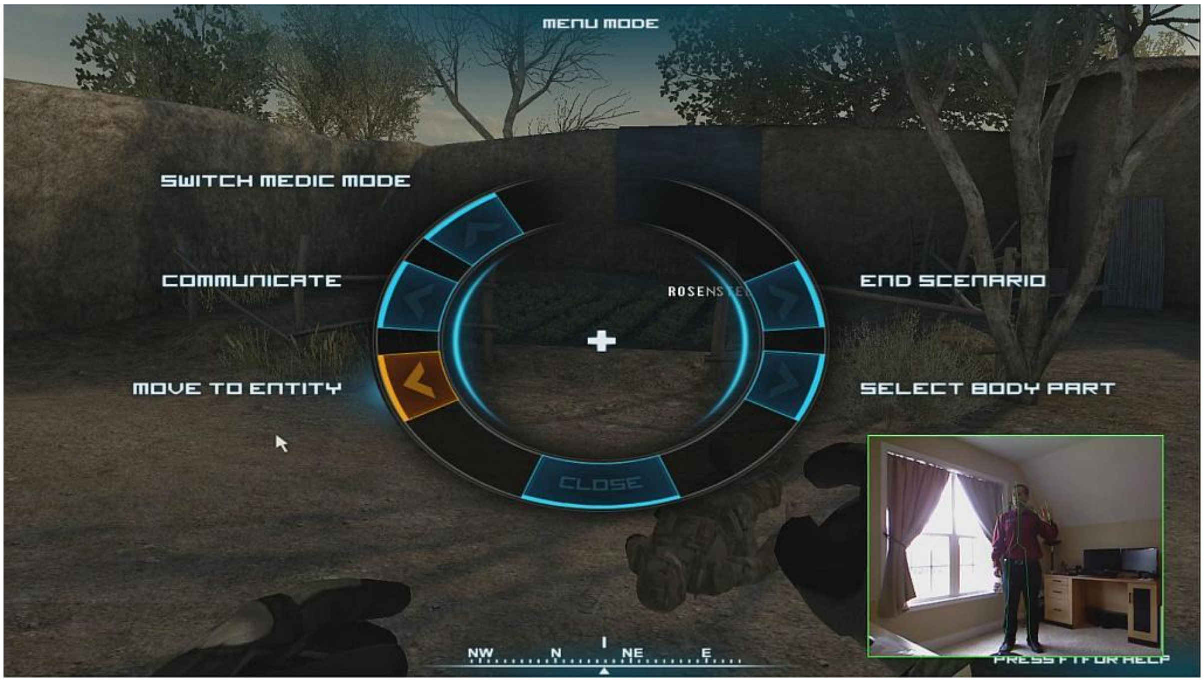

Figure 1: User interacting with menu-driven interface

This approach had some problems. The navigation of the user's point of view and position in space proved to be imprecise. The user would often move past their goal and then have to make corrections; over-control was a huge issue. It took a long time for a user to select a body part using this technique and required a large amount of tedious concentration. It was frustrating and difficult to get the menu to pop-up. Using the Kinect with the original TC3Sim paradigm also had another issue. Because the interface relied on a high-degree of a precision provided by a mouse, many of the gestures became difficult to execute. One prime example is selecting a menu option for the contextual menu. A user would struggle to get the correct menu to pop up, and then would tediously place the cursor on the button of their selection. Upon raising their other hand to click, their body would shift and the cursor would move, resulting in the selection of a different option than intended or no selection at all. This proved frustrating for the developers testing the interface they had just built and it was clear (even without a larger usability test) that this would not work.

The interface needed to be re-thought. Greater tolerance for menu selections corrected for the lack of precision inherent in the Kinect interface. Confusion caused by the system's inability to distinguish similar gestures generated the concept of contextual modes. Lastly, simpler gestures closely mimicking actual medical procedural movements simplified certain in-game actions.

Different distances proximal to the screen created different contextual modes in TC3Sim for Kinect. There were three zones. The furthest is a NUI inactive zone. The middle distance is for navigation. The closest distance is for treating a casualty. The contextual zones are approximately one large step apart. The user enters any zone by stepping forward or backward. Upon entering any zone, an audio cue indicates that you have made the transition. The border of the video feed in the lower right corner of the screen also changes to indicate which mode the user is in.

The furthest zone is a NUI inactive zone. This allows the user to start the system or use a game controller without inadvertently causing the NUI to react. Having a safe haven allows the user to have someplace to relax without having to worry about their body.

The middle distance encompasses the navigation context. The user can navigate the environment by pointing their shoulders as in the original NUI implementation. Moving a single leg forward and backward allows walking in the virtual environment. The user interface allows a user to move their legs without causing a transition to another mode. Modes are based on whole body position. Plus, a little lag is built into the system when determining a zone change to avoid false positives. A user can still access the contextual menu in this mode by lifting one hand. The menu was redesigned so a selection is based on a whole quadrant of the screen. To aid in navigation, a menu option was added that automatically takes the user to a selected casualty. This avoids many of the over-control issues that were prevalent in the first implementation.

The closest distance allows for interaction with a casualty. Simple gestures are used in treatment rather than menu selections. When the player places their hand on their forehead, TC3Sim focuses on the casualty and their main injury. For amputation treatment, the player must apply a tourniquet. To do this, the player places their hand on either their shoulder or their elbow. A shoulder indicates applying the tourniquet at the upper part of the limb, and an elbow indicates applying the tourniquet lower. This gesture needs to be on the same side as the amputation. When the player waves their hand up and down, a blood sweep is performed on the casualty. When the player waves their hand from side to side, TC3Sim applies a bandage; or if there is already a bandage present, removes it (Figure 2).

Figure 2: User utilizing gestural interface schema

These combinations of improvements in initial testing by the developers proved much easier and more natural. There are three guidelines that came out of the redesign:

1. If the hand must select something, make sure that something is big,
2. Allow for a safe zone where the player can relax without fear of unintended consequences, and
3. In a complex system with a large number of interactions, context needs to be utilized to simplify the interface for the user and to allow it to be perceived as more natural.

USABILITY METHODOLOGY

To get initial data on the system for further refinement, the researchers conducted an initial usability study. The study included a total of five participants, primarily West Point Cadets. Studies have shown that the first five participants uncover the majority of usability issues (Lewis, 1994; Neilsen and Landauer, 1993; Virzi, 1992). For the purposes of an iterative usability study such as this, the low number of participants is ideal for identifying major design faults while minimizing resources that would be required for a large scale evaluation. During the study, the participants had one to two minutes to review the instructions for the required tasks. The instructions detailed the hand and body movements associated with making the different gestures, moving throughout the terrain, and navigating the menu screens. The instructions also identified six key tasks for the user to perform: focus on the wound, apply a tourniquet, perform a blood sweep, apply a bandage, request a security sweep, and request a medical evacuation. In addition to the written instructions, a laboratory assistant demonstrated the necessary body movements to each participant.

Once participants finished the instructional period, the laboratory assistant initialized the simulation and verified that the Kinect successfully recognized each participant without obstructions from background movement. After the Kinect calibration, participants entered into the virtual environment. The scenario required participants to locate the casualty, approach the casualty for treatment, treat a lower arm amputation, and request a security sweep and medical evacuation. The laboratory assistant monitored each participant's progress within the scenario, evaluating task success using a checklist. In addition, the laboratory assistant used a stopwatch to record the time each participant took to apply a tourniquet and the total time to complete all of the required tasks.

Upon completion of the scenario, each participant received a 22 question survey to evaluate their experience during the study. The participants' responses served as their personal assessment of the user interface. Participants rated each question on a 5-point Likert scale ranging from strongly disagree (1) to strongly agree (5). Participants also provided open ended comments on the surveys to address issues not specified in the survey questions.

USABILITY RESULTS

All participants successfully completed the scenario by requesting the medical evacuation. Additionally, all participants applied the tourniquet successfully, which represents the most important training objective and the key life saving medical intervention in the scenario. All participants also conducted a blood sweep, ensuring there were no additional injuries requiring treatment. Only 60% of participants completed the bandage application, indicating a potential issue with that task. Lastly, 80% of participants completed the security sweep, with one participant forgetting to complete the task (Figure 3).

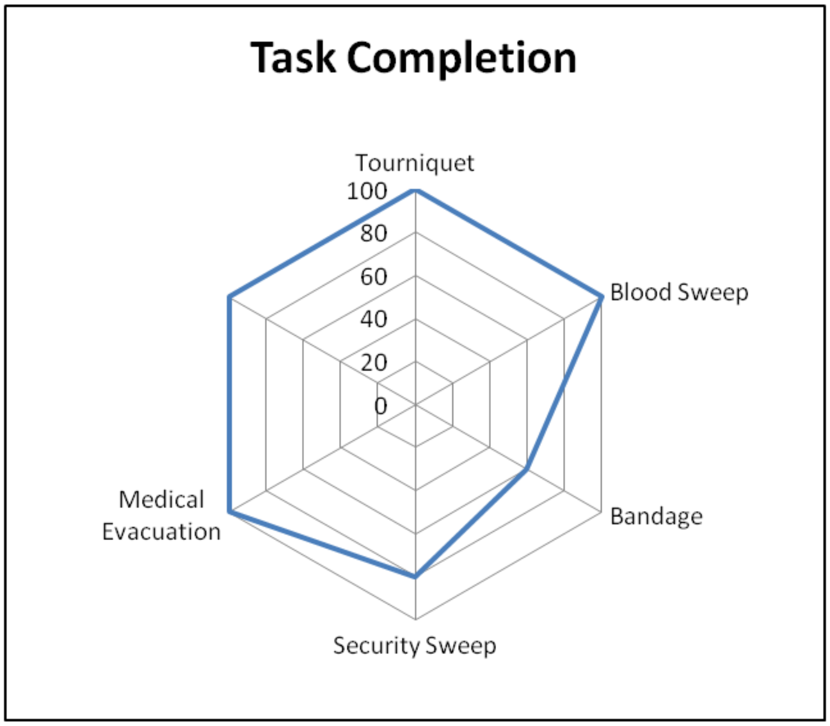

Figure 3: Task Completion Radar Plot (percentage of participants)

Participants completed the tasks in a time frame which would be acceptable for training purposes. The average time to place the tourniquet was 117.2s, and the average total time was 265.2s (Figure 4). The times varied widely, with some participants taking time to explore the environment and menu options.

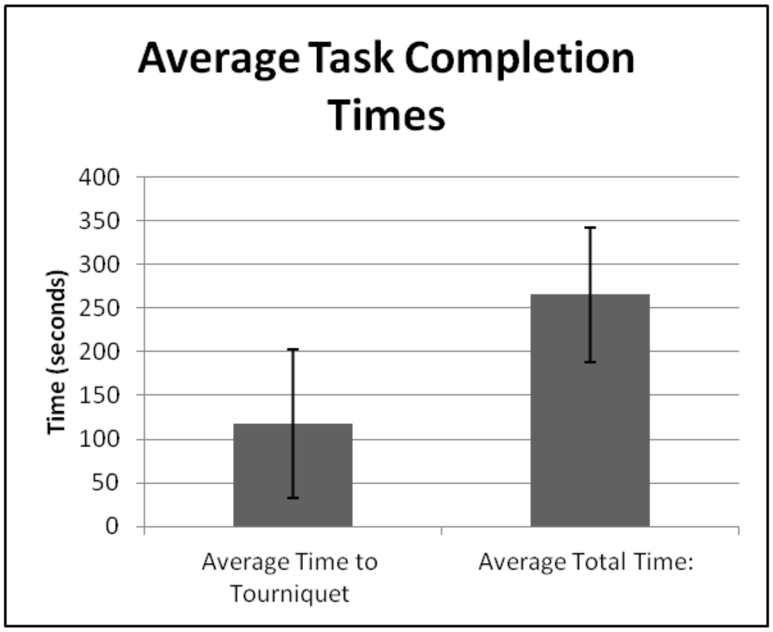

Figure 4: Average Completion Times (95% CI Shown)

The survey results indicated both positive and negative perceptions within the simulation. To begin, the overall perception of the system was average, with a mean score of 3.12 with a confidence interval of ±0.67. The gestures received high ratings from the participants, with a mean score of 3.733±0.86. The menu system received varied results, with users on both ends of the spectrum, and the results were a mean score of 3±1.3. The most user frustration and subsequently the lowest user ratings was related to virtual locomotion using the Kinect, with a mean score of 2.5±0.96. With the locomotion scores removed, the overall simulation score was significantly improved, with a mean of 3.46±0.66 (Figure 5).

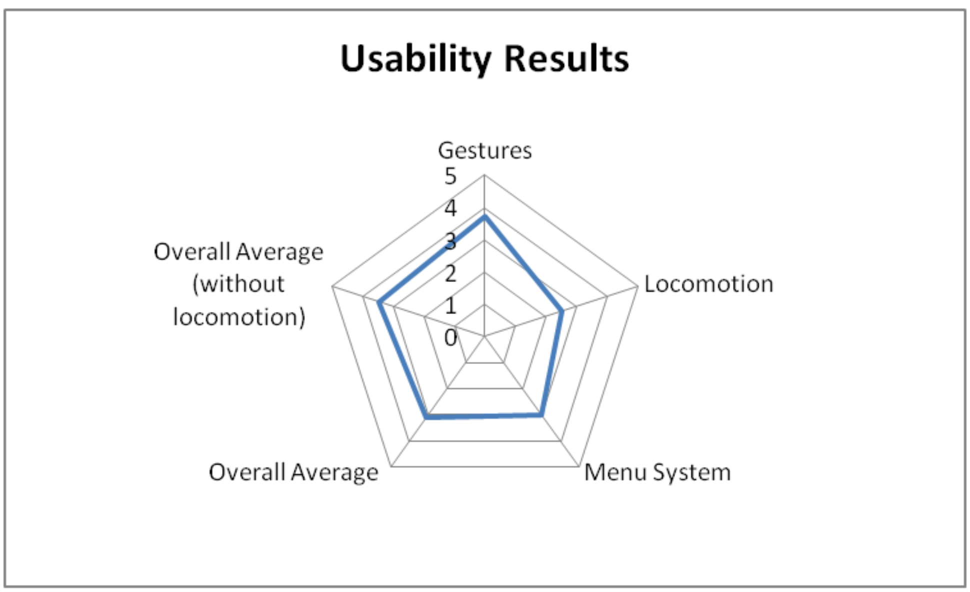

Figure 5: Usability Results Radar Plot

CONCLUSIONS

The usability results indicate that while users liked portions of the interface implementation, other areas represented significant frustration. The primary area of frustration related to virtual locomotion using the Kinect. Users had difficulty moving and looking around in the environment. For example, users attempting to look up had to bend backward awkwardly at the waist. Researchers also noted user frustration caused by the delay from the Kinect when looking around. Many users turned too far requiring them to re-adjust. These frustrations are evident in other studies as well, which note a difficulty tracking the legs and feet during locomotion activities (Williamson et al., 2011). Williamson sought to overcome this challenge using multiple Kinect sensors, but that effort is still in its infancy (Williamson et al., 2012). The menu system also caused frustration in users primarily due to inadvertent hand motion. Despite the large menu buttons, the Kinect sensor was sensitive enough to track small movements of the hand, which resulted in users being unable to select menu items before the cursor moved. Further frustration resulted from a delayed button click mechanism. The initial design intended to minimize accidental button clicks through this delay, but the delay also caused frustration. The initial internal reviews also noted these issues, which led to the majority of the redesigns. The area with the most positive user responses was the gestural interface. These gestures increased the level of immersion within the scenario by requiring users to replicate motions similar to those made during treatment. Furthermore, since user movements are similar to actual treatment movements, the interface proved to be more intuitive. The intuitive nature of the gestural interface resulted in a high level of user satisfaction. The majority of users successfully executed all the gestures, indicating that they were simple to perform. The success users experienced with the gesture system contributed to user satisfaction. The results of this research effort indicate that a gestural interface system is well-received but requires careful attention to the interface design. The usability study results generated the following guidelines:

1. Simple gestures which are closely tied to real-world movements are excellent control mechanics
2. Locomotion can easily become a frustration in a virtual environment; adding shortcut mechanics to focus on areas of interests circumvents many of the frustrations
3. Delaying the click mechanism to remove false positives is frustrating when the user perceives a significant lag effect
4. Visual and auditory feedback is a necessary indicator for users to determine successful gesture or movement

The Kinect sensor represents a very low-cost solution for gestural interfaces. While the current implementation of the Kinect interface is lacking for virtual locomotion and menu driven applications, the gestural capabilities are significant and are well-received by potential users. With an array of novel sensors emerging, including the next generation Kinect, the Leap motion sensor, and the Myo arm sensor, the notion of a truly intuitive interface may be on the horizon. If none of these sensors are capable of meeting all the desired interface goals individually, sensor fusion may achieve these goals by networking a variety of sensors together. Further research is necessary in this area before a solution for a truly immersive, natural user interface is within reach.

REFERENCES

Butler, F. K. (2010). Tactical combat casualty care: update 2009. The Journal of Trauma and Acute Care Surgery, 69(1), S10-S13.

Lowensohn, J. (2011). Timeline: A look back at Kinect's history. Retrieved May 24, 2013 from http://news.cnet.com/8301-10805_3-20035039-75.html.

Mania, K., & Chalmers, A. (2001). The effects of levels of immersion on memory and presence in virtual environments: A reality centered approach. *CyberPsychology & Behavior*, *4*(2), 247-264.

Solstad, J (2011). Evaluating the effect of downtime on satisfaction levels of military training programs. *National Council for Undergraduate Research*.

Sotomayor, T. M. (2010). Teaching tactical combat casualty care using the TC3 sim gamebased simulation: a study to measure training effectiveness. *Stud. Health Technol. Inform*, *154*, 176-179.

Sotomayor, T. M., Quintero, B., Salva, A. M., & Sherrill, J. (2012, January). High Fidelity Physiological Model for Immersive Simulation and Training. In *The Interservice/Industry Training, Simulation & Education Conference (I/ITSEC)* (Vol. 2012, No. 1). National Training Systems Association.

Williamson, B. M., Wingrave, C., LaViola, J. J., Roberts, T., & Garrity, P. (2011, January). Natural Full Body Interaction for Navigation in Dismounted Soldier Training. In The Interservice/Industry Training, Simulation & Education Conference (I/ITSEC) (Vol. 2011, No. 1). National Training Systems Association.

Williamson, B. M., LaViola, J. J., Roberts, T., & Garrity, P. (2012, January). Multi-Kinect Tracking for Dismounted Soldier Training. In The Interservice/Industry Training, Simulation & Education Conference (I/ITSEC) (Vol. 2012, No. 1). National Training Systems Association.