Evaluating the Potential of Cognitive Rehabilitation with Mixed Reality

Nicholas Beato, Daniel P. Mapes, Charles E. Hughes,
Cali Fidopiastis, and Eileen Smith

University of Central Florida
Orlando, FL 32816
nbeato@eecs.ucf.edu, dmapes@ist.ucf.edu,
ceh@eecs.ucf.edu, cfidopia@ist.ucf.edu, esmith@ist.ucf.edu

Abstract. We describe the development and use of a mixed reality (MR) testbed to evaluate potential scenarios that may alleviate performance deficits in subjects who may be experiencing cognitive deficiencies, such as posttraumatic stress disorder (PTSD). The system blends real world sensory data with synthetic enhancements in the visual and aural domains. It captures user actions (movement, view direction, environment interaction, and task performance) and psychophysical states (engagement, workload, and skin conductivity) during an MR-enabled experience in order to determine task performance in the context of a variety of stimuli (visual and aural distracters in time-constrained activities). The goal is to discover triggers that affect stress levels and task performance in order to develop individualized plans for personal improvement.

Keywords: Mixed reality, post traumatic stress disorder, psychophysical sensing, medical rehabilitation, cognitive rehabilitation.

1 Introduction

In cognitive rehabilitation, it is vital that a patient’s training generalizes and transfers to everyday situations. Environmental and economical factors have motivated medical researchers to explore alternatives including virtual environments in order to provide safe, reusable, and cost-effective recovery [4]. A drawback to virtual reality (VR) technology is that a patient only interacts with two senses, sight and sound, which the system overrides with artificial content. Seemingly normal expectations, such as manipulating physical objects with your hands, may prevent the patient from fully investing in the experience, compromising treatment. Mixed reality (MR) technology aims to seamlessly and believably combine virtual and real world content in a safe and controlled setting. To apply such technology to rehabilitation of cognitive disorders, it is necessary to evaluate treatment scenarios in a safe, cost-effective, and non-subjective manner. Our MR toolkit augments the visual and auditory senses while allowing the other senses to draw from the actual environment. With the addition of psychophysical sensors, we hypothesize that monitoring healthy participants may aid the efficacy of particular environments for the treatment of cognitive disorders [3], [4].

R. Shumaker (Ed.): Virtual and Mixed Reality, LNCS 5622, pp. 522–531, 2009.

2 Mixed Reality: What, Why, How?

2.1 What Is MR?

As defined by Milgram et al. [6], MR covers the spectrum of experiences ranging from almost purely virtual to almost purely physical. That is, an MR experience must involve some virtual and some real world aspects, with the amount of each highly dependent upon the application's requirements. Real experiences with added virtual overlays are categorized as augmented reality (AR). Synthetic experiences with added real world components are categorized as augmented virtuality (AV). In the case of AR, we generally overlay virtual objects on top of real ones, e.g., adding identifiers like textual or iconic information that make a visual landscape more understandable. In the case of AV, we typically place a small interactive set in a virtual surround, e.g., the user sits on a real chair at a real table interacting with other people within the context of a virtual restaurant.

The most interesting and challenging part of the MR spectrum lies in the middle, where virtual and real objects coexist. In order for these objects to coexist, they must interact with each other based on the user's viewpoint. Obviously, it's trivial for real objects to coexist with other real objects, and virtual objects to coexist with other virtual objects. The challenge lies in the interaction between the real and virtual.

2.2 Why MR versus VR?

VR can be viewed as the extreme limit of MR in which all assets are synthetic. A primary attribute of VR is that it dominates users' senses, separating them from the real world in order to provide a purely synthetic experience. While such isolation from reality is useful in some applications, it is very limiting in others, especially those in which people are performing tasks in a context that involves all their senses or relies on triggering memories in order to interact with objects and people [4]. In effect, VR takes its users away from the physical context, whereas MR enhances the physical context. This is a very important distinction for cognitive assessment and rehabilitation, where we often want to determine and address a subject's deficiencies at performing tasks in a specific, realistic context.

VR tends to be a visual and aural experience, with other senses only peripherally addressed. Because MR can take advantage of the real world, it usually involves all senses, with the visual and aural being a blend of real and synthetic, and the other senses generally being real. For instance, a rehabilitation experience could take place in a MR replica of a patient's kitchen [3], where the counters, the cabinets and their contents, a refrigerator, a toaster oven, a coffee maker and an accompanying therapist are real. These real assets are augmented by virtual textures on the counters and cabinets. The scene might also include a virtual wall with a virtual window and its accompanying virtual landscaping, a virtual eating area, and optional virtual aural cues to help the patient carry out tasks in the correct sequence. Passive and active haptics are provided by the environment's real objects, e.g., you can lean against the counters while opening the cabinets. The smell of food, e.g., a bagel toasting, and its taste are real. This combination of real and virtual content makes for a rich experience that may trigger old memories for patient assessment and build multiple pathways to new and existing ones for rehabilitation.

2.3 How Can We Achieve MR?

The most recognizable aspect of a functioning MR system is the ability to trick the brain into perceiving the presence of things that are not really there. Visually, we must enhance the appearance of the real world from the user's point-of-view. While other techniques exist, a see-through head-mounted display (ST-HMD) is a common way to accomplish this. We specifically use a video ST-HMD (VST-HMD), a device that allows us to capture the user's view, in stereo, from optically aligned mounted cameras, process the video on a computer system in real-time, and display the composited result, properly registered, via small LCDs in front of the user's eyes. Because the system acquires video and augments it before the user perceives the real content (as opposed to rendering on a transparent screen, as done on an optical ST-HMD in AR), we have much more control over the registration and synchronization of the real and virtual worlds. This control is a necessity, as the experiences we develop often require multiple, alternating layers of real and virtual content.

Although audio is often deferred to the end of production in the development of multimedia experiences, it is integral to a person’s experiences and their recollection. Thus, for any MR experience intended for assessment or rehabilitation or for that matter nearly any interactive experience, the aural components are as important as the visual, and should be part of the entire design process, not a last-minute add-on. Just as the visual component, the aural part of MR involves many challenges. The challenges are understanding the sources of real audio, especially when there are multiple, concurrent origins; blending the real and virtual; and delivering this blend into a complex landscape [5].

To properly register the virtual content, whether visual or aural, in relation to the real, the underlying MR software needs to know the user’s head position and orientation. Detecting an object’s 3D position and orientation in physical space is commonly referred to as six degrees of freedom (6DOF) tracking. Until recently, accurate 6DOF tracking of people and objects in an MR setting was a costly proposition, involving expensive magnetic, acoustical, optical and/or inertial systems. In contrast, newer technology, such as infrared cameras, now provides the basis for very inexpensive yet accurate tracking. Advances that must still be made are in the vision algorithms for detecting and differentiating markers, and for recognizing gestures in order to provide semantic interpretation to people’s non-verbal communication. Having an understanding of the meaning of movements allows one to use the human body as an interface device, to easily compress communications of actions for networked multiplayer experiences and to drive a simulation based on a user’s body language.

In order to capture a reasonably complete picture of an experience, MR systems must continuously record a user’s movement, view direction, interaction with environment and task performance. This must all be correlated with the participant’s psychophysical states, which can be monitored through unobtrusive, wireless biosensors (EEG, ECG and respiratory, temperature and electrodermal) [1]. EEG measures of engagement and workload can assist in determining the efficacy of a MR based rehabilitation environment within the feasibility stage [2]. These combined data recordings can be used to determine how well subjects perform tasks in the context of a variety of stimuli (visual and aural distracters, and time-constrained activities). Most importantly, the captured data can be visualized and then used by therapists to understand the patient's unique condition.

In order for commercialization to be an eventual outcome of a research project, one must address the important criteria of reliability, scalability and cost. Reliability is approached primarily through the use of commercial off-the-shelf hardware and strong software engineering practices. Scalability is insured by designing experiences that can be delivered in a tiered fashion, ranging from a VST-HMD to a full surround (circular or four-wall) stereo to a single wall stereo or mono version. Cost is addressed through the use of commodity hardware, free-license software, vision-based tracking, and the development of carefully crafted stories as a way to deliver contextually meaningful experiences with commodity hardware.

3 Software Infrastructure

3.1 DNA Engine

To facilitate agile development of low-cost, reliable and scalable scenarios ranging from VR to AR, the Media Convergence Lab (MCL) is iteratively developing a component-based engine, dubbed DNA, which loads and configures seemingly complex objects from reusable modules via XML. Modules may provide direct interfacing to available pre-existing libraries or may be user-friendly proxies to ongoing research code. The main development goal is to allow non-programmers to quickly assemble experiences using examples as templates. In other words, we note that project code is mostly "copy/paste" code that is thrown together, so we encourage this process by providing working examples for different aspects of the system and keeping this methodology in data files rather than source files when possible.

To allow rapid prototyping of new features, MCL has chosen to find suitable open-source or free-license libraries and progressively expose these libraries on an as-needed basis to the DNA loader as plug-ins. Such bindings can typically be done by directly mapping necessary library objects to XML elements, where public data members have XML attributes with identical naming conventions. This strategy allows us to point scenario developers to existing library documentation when such documentation is available.

For MR experiences, the DNA engine has several components that we repeatedly employ. On the display side of things, we expose Object-Oriented Graphics Rendering Engine (OGRE) for modern graphics presentation, which is capable of supporting VST-HMDs without any additional source code, and Cross-Platform Audio Creation Tool (XACT) for audio delivery, which enables custom source-to-speaker attenuation for non-conventional speaker configurations. In addition to synthetic contributions, we also have controls for digital multiplexing (DMX) to manipulate most powered real-world devices. On the input side of things, we have access to the basic keyboard and mouse though Object-Oriented Input System (OIS), but we also provide modules for a large number of devices including the WiiMote, 6DOF tracking, and bar code scanners. There are also a few hybrid libraries providing features such as multi-touch surface support.

A particularly useful feature of the DNA engine is that all data is exposed in the XML document. By cleverly nesting elements, we can quickly find pertinent information in another component, allowing specialized code to either subscribe as a data consumer or publish as a data provider. This dataflow approach decouples source from unnecessary dependencies and allows easy modification for both debugging and project changes. A particularly useful application of this feature arises during after-action review (AAR). For example, since we are interested in logging the position of a user for later playback, we can subscribe to the element that represents the user's position in the document and log changes observed in the element's attributes over time. Furthermore, the changes can occur from another, unknown element, such as a 6DOF tracker, network device, or keyboard that interprets values acquired from hardware. To construct an AAR tool for an experiment, we replace XML elements that interactively update the system state with proxies capable of loading recorded data files and mimicking the interactive modules. With other additions, such as the ability to manipulate system time and visual enhancements of important events, we can prototype project-specific AAR systems in a short amount of time and add metrics as researchers require them.

3.2 Blue-Screening

Chroma-keying, commonly referred to as blue-screening, is a viable method to inject virtual objects into the background of a video. Most chroma-keying tools target ideal cameras and lighting when applied in real-time, such as live news broadcasts. In MR, this problem is complicated by the sheer amount of processor time required by other aspects of the system and the relatively poor quality VST-HMD cameras. To alleviate these concerns and apply chroma-keying, we developed a GPU-based algorithm to detect and modify blue (or whatever color is used in the background) pixels.

The primary goal of our chroma-keying method is to provide a basic semi-automated interface for quick calibration in controlled environments that may be performed by non-experts. To realize this goal, we note that a camera digitizes the physical chroma-key material in such environments as a near solid color. The training step requires a user to view the surrounding material, allowing the system to statistically analyze the color and produce a parameterized, iterative fragment program capable of determining a good estimate of the opacity of each pixel in parallel on a GPU. This step may be done by a participant during the scenario, but is better suited for a setup procedure. There are two optional parameters exposed for tweaking that indicate confidence intervals, in standard deviations, of the training data. These parameters are necessary in ill-conditioned setups, such as situations in conference hall lighting and experiences with noisy VST-HMD cameras. Otherwise, the system can interactively determine the alpha matte from a video for virtual scene compositing with no user input aside from the initial training capture.

We note that under our camera restrictions and delivery method, dependent on noisy video input and real-time stereo rendering, we tend to aggressively key pixels and attempt error correction by adapting foreground extraction techniques. This methodology is not well-suited for off-line processing of high quality video, as it results in some misclassifications, especially in shadows and around objects, than what is commonly seen and accepted in the industry. Advances in VST-HMD, MR, and image matting technology will account for these downfalls in time. Our focus is to allow quick calibration in the field and deliver a believable experience to the user.

4 The MR Warehouse

For a first phase project, our goal is to develop a system capable of recording data that may eventually aid in the assessment and treatment of cognitive disorders, such as posttraumatic stress disorder (PTSD). As a feasibility test, we developed a MR scenario that puts a healthy participant into the role of a warehouse employee using a Canon COASTAR VST-HMD [7]. Our goal is to capture data to determine whether or not post-session analysis can evaluate the scenario's efficacy. For each test, the participant is responsible for several tasks, including buzzing in delivery trucks, fulfilling printed orders, and keeping track of inventory. Baseline tasks are performed in a quiet room to calibrate the sensors, and then they run through the full-fledged MR system. The system monitors both the user actions (gaze direction, movement, task performance, and environment interaction) and physiological state (EEG, ECG and respiratory, temperature, and electrodermal). This provides several metrics (such as engagement, workload, and skin conductivity) capable of correlating the physical state of a participant to task performance. In order for these metrics to have useful meaning, we must make the scenario interactive and realistic. This allows us to safely determine the difficulty of a set of tasks for an impaired subject by validating and inferring from test run on healthy people.

To provide tactical response to the virtual environment, we utilize chroma-key blue paint for the work surface and large cabinet. The virtual environment contributes to the appearance of these objects.

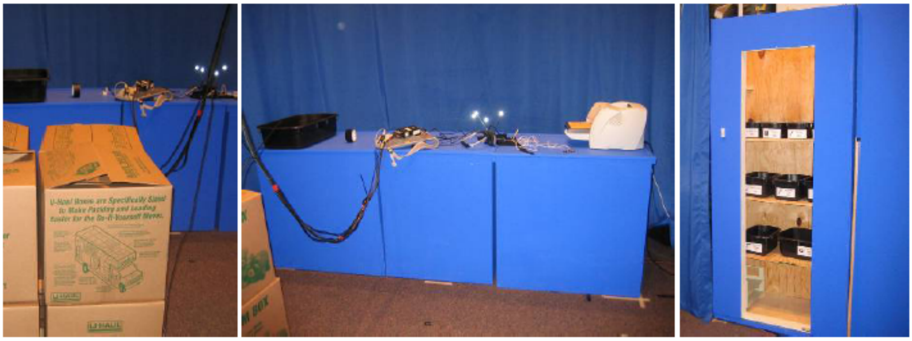

Fig. 1. In the physical environment, we have a combination of chroma-key blue material objects and visible objects. The chroma-key allows overriding the appearance of real objects, enabling the patient to believably touch virtual surfaces.

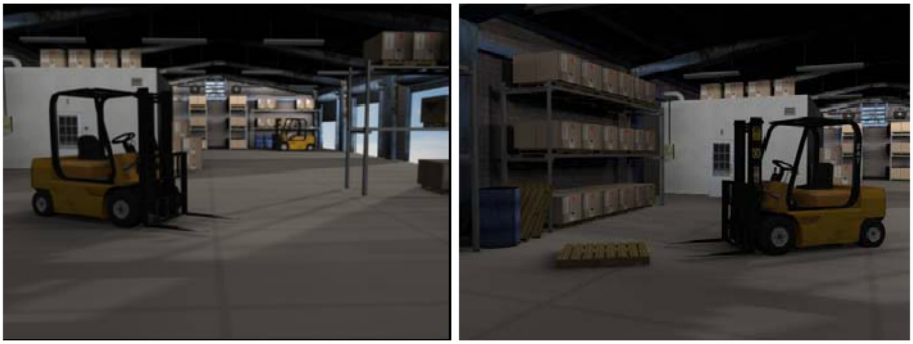

Fig. 2. The hazardous environment is purely virtual (including audio). This allows us to cheaply and safely introduce distracters into the rehabilitation scenario.

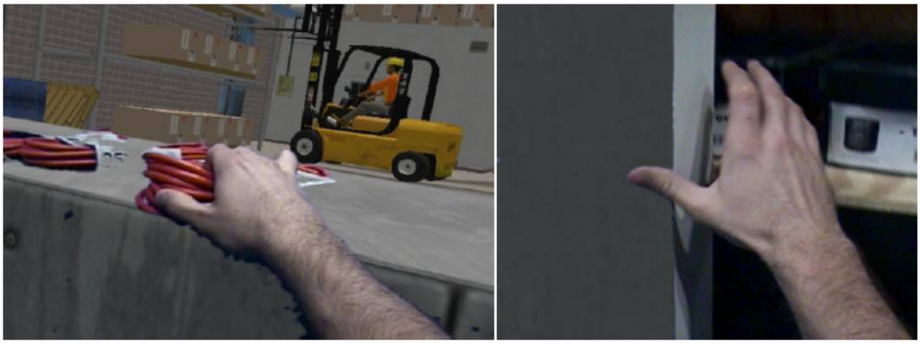

Fig. 3. The MR world allows the participant to see, hear, and feel a blending of the real and virtual worlds

We also use real objects as interactive props, such as the printer, buzzer, scanner, and inventory items, for order fulfillment and inventory stocking (Figure 1).

In the MR experience, the user can physically interact with these objects as one would in the real world, something that is infeasible in VR right now.

The real environment is then augmented with 3D graphics and spatial audio. These extensions safely immerse the subject in an otherwise hazardous environment, where moving forklifts, delivery trucks, and heavy boxes could cause complications or injury (Figure 2 and Figure 3). In addition, the physical setting, including both aural and visual, primarily occurs in the virtual world, allowing the scenario to restart (or run multiple times) with substantially less effort for a new participant, thus reducing the cost of subsequent evaluations. This also aids the portability of the experience.

Tracking user's interaction with the environment poses some difficult questions. To address these concerns, we use story as a seemingly natural interface to the simulation system. In particular, there are two types of tasks we must detect: item tracking and environmental awareness. Each time a participant focuses on a real item and moves it, the system should reliably react. To accomplish this, we require the participant to scan the item, taken from a known location, and to sort it by both picture and name in the destination storage cabinet. While this might seem unnecessarily complex, creative use of story and experiment design allows these steps to seem like a natural part of the task. Specifically, we instruct the user that filling an order, as a job requirement, includes the responsibility of tracking each item with a barcode scanner. Similarly, automated detection of the user's awareness of incoming delivery trucks (which is primarily heard in audio, but is also visible with the opening and closing of delivery bay doors) requires that the user presses a button to notify the manager. Through the use of story and context, this input seems natural, if not necessary, to the user. In reality, it gives us a means to track the user's situational awareness when trucks arrive in the scenario. In both situations, we motivate the user to provide us with systematic, reliable feedback by using story to reinforce that the performed action is both necessary and reasonable.

5 Experimental Findings

Since only five of twelve healthy individuals who participated in our study provided reliable data (others recorded too much noise during aspects of the scenario), we cannot draw any general conclusions. However, we can note common observations. Overall, participants showed a mix of both high and low engagement with frequent distractions. Distraction was mostly associated with the audio stimuli for the printer. Spikes ranging from 50 to 70% distraction were classified within 10 seconds of one or more printer audio cues. This quantifiably indicates that a particular scenario event had a negative effect on the participants without the use of post-test questionnaires, allowing us to make alterations to the environment.

When looking at workload through the scenario, the participants show phases of high and low levels. The changes in workload are not correlated with any particular aspect of the scenario. Further analysis of individual performance including task strategy may show distinct associations between tasks that demand more cognitive resources than others. We also observed that users only buzzed the manager when they visibly saw the first delivery truck. The average high workload prior to the arrival of all trucks is relatively high (near 60%), with the remainder dominated by moderate workload. For reference, this is enough to burn out a healthy employee within a week. From this, we can tell that participants were too focused on the current task, to the point that environmental awareness was compromised. This could explain why the sound of trucks was missed. However, we can also assume that the audio cues were not prevalent enough in the simulation and adjust settings accordingly. The important point is that we know that we overloaded healthy patients. So, we can detect this during the experiments provided we have a baseline to compare against.

Overall, the results show that the EEG measures of engagement and workload are good indicators of how the tasks affected the healthy participants. This data can be analyzed individually and in aggregate to obtain an understanding of cognitive aspects of the tasks that may pose challenges to head injured patients. For example, we can safely determine that a task is too hard for a healthy human, concluding that it would frustrate a rehabilitation patient with potential negative consequences. Furthermore, we can detect this during an experiment. This information is imperative to know for not only virtual rehabilitation therapy protocols, but also the field of rehabilitation in general.

6 Conclusions and Future Work

The primary goal of MR-enabled treatment is to alleviate a subject's deficits as regards task performance in real-world contexts. We have demonstrated that a combination of psychophysical and simulation-oriented data metrics promises to provide useful indicators of the effects of scenarios on participants. Specifically, we found that pre-experience calibrations and experience-time data capture allowed us to assess the stress level changes in healthy subjects in the context of task performance and simulation events. This provides encouraging indications that MR can be used in the assessment of affected populations and that the results of these MR-enabled patient assessments might be used to create therapeutic plan. Once such a plan is developed, traditional and/or MR-enabled therapy can be applied. In the case of MR-enabled therapies, we note that the system described here allows a therapist or technician to modify such a plan during run-time, potentially improving the course of treatment.

Based on our current findings, our next step is to apply these MR techniques to an affected population. If the outcomes are successful in isolating triggers that adversely affect performance of affected individuals, as they were with the unaffected population, then we will proceed to our primary goal, that of applying MR during the rehabilitation phase, using it to improve performance and, where appropriate, stimulate and enhance cognitive functions and induce positive neuroplastic changes.

Acknowledgments

The research presented here is partially supported by the National Science Foundation (IIP0750551 and DRL0638977) and the Air Force Office of Scientific Research.

References

1. Berka, C., Levendowski, D.J., Cvetinovic, M., Petrovic, M.M., Davis, G.F., Lumicao, M.N., Popovic, M.V., Zivkovic, V.T., Olmstead, R.E., Westbrook, P.: Real-time Analysis of EEG Indices of Alertness, Cognition and Memory Acquired with a Wireless EEG Headset. Special Issue of the International Journal of Human-Computer Interaction on Augmented Cognition 17(2), 151–170 (2004)
2. Fidopiastis, C.M., Hughes, C.E., Smith, E.M., Nicholson, D.M.: Assessing Virtual Rehabilitation Design with Biophysiological Metrics. In: Proceedings of Virtual Rehabilitation, p. 86. IEEE, Venice, Italy (2007)
3. Fidopiastis, C.M., Stapleton, C.B., Whiteside, J.D., Hughes, C.E., Fiore, S.M., Martin, G.A., Rolland, J.P., Smith, E.M.: Human Experience Modeler: Context Driven Cognitive Retraining to Facilitate Transfer of Training. CyberPsychology and Behavior 9(2), 183–187 (2006)

4. Fidopiastis, C.M., Weiderhold, M.: Mindscape Retuning and Brain Reorganization with Hybrid Universes: The Future of Virtual Rehabilitation. In: Schmorrow, D., Cohn, J., Nicholson, D. (eds.) The PSI Handbook of Virtual Environments for Training & Education: Developments for the Military and Beyond, vol. 3, pp. 427–434. Praeger Security International, Westport, CT (2008)
5. Hughes, D.E.: Defining an Audio Pipeline for Mixed Reality. In: Proceedings of Human Computer Interfaces International, Lawrence Erlbaum Assoc., Las Vegas (2005)
6. Milgram, P., Takemura, H., Utsumi, A., Kishino, F.: Augmented Reality: A Class of Displays on the Reality-Virtuality Continuum. In: Telemanipulator and Telepresence Technologies, vol. 2351, pp. 282–292 (1994)
7. Uchiyama, S., Takemoto, K., Satoh, K., Yamamoto, H., Tamura, H.: MR Platform: A Basic Body on Which Mixed Reality Applications Are Built. In: IEEE and ACM International Symposium on Mixed and Augmented Reality, pp. 246–256. IEEE Computer Society Press, Darmstadt, Germany (2002)