The Effects of Fully Immersive Virtual Reality on the Learning of Physical Tasks

Kayur Patel
University of Washington
{kayur@cs.washington.edu}

Jeremy N. Bailenson
Stanford University
{baileson@stanford.edu}

Sang-Hack Jung, Rosen Diankov, Ruzena Bajcsy
University of California at Berkeley
[sangj,ffxvross,bajcsy]@eecs.berkeley.edu

Abstract

Fully immersive virtual settings are different from traditional virtual reality settings in that they are able to capture full body motion. This ability allows people to use their full range of physical motion to interact with other avatars, computer controlled agents, and objects in the virtual environment. As such, fully immersive virtual reality presents a novel mediated learning environment in which people can learn physical activities. Capturing human motion for virtual settings has traditionally been a model-based approach where a few degrees (on the order of tens) of freedom are mapped to virtual model. In contrast, we use an image-based solution that sacrifices visual fidelity for motion fidelity and increased degrees of freedom (on the order of hundreds). Due to the difficulties involved with building such an image-based immersive system, very little work has been done to assess the effectiveness of this form of mediated learning. In the current work, participants were taught several tai chi moves in either a 2D video system or a 3D immersive system equipped with features not possible to implement in traditional video systems. We demonstrated via blind coder ratings that people learned more in the immersive virtual reality system than in the 2D video system, and via self-report ratings the social presence was higher as well. We discuss these findings and the resulting implications for designing and testing fully immersive systems.

Keywords--- **virtual reality, human computer interaction, mediated learning, computer vision**.

1. Introduction

From entering a holodeck on a starship to plugging into the matrix, our culture has been fascinated with virtual settings, indistinguishable from reality, which we can augment at our whim. In the pursuit of this ideal, researchers have developed various virtual reality (VR) systems. Creating high fidelity virtual environments is a difficult endeavor; no current system has adequately solved this problem. Nevertheless, as systems have improved, VR has proven useful for various domains; in particular, it has been shown to be a successful environment for learning [1].

Research has shown that, for learning physical motion, repeating a task reinforces learning [2]. Virtual reality, as a platform for mediated learning, provides instruction on-demand, allowing students to repeat difficult motions whenever instruction is needed and as often as is needed away from the social pressures of a classroom setting. Learning adapts to the user's schedule, goals, and speed. Previous research has shown the affordance of on-demand learning provides a distinct advantage over face-to-face human interaction [3].

While some virtual systems have proven successful for learning, most VR systems lack the ability to capture the user's full range of motion, limiting their ability to fully immerse the user in the virtual setting. In contrast, fully immersive virtual reality allows for full mobility by capturing human motion and reproducing the same motion in the virtual representation. Full immersion is crucial to any type of learning activity that involves any type of body coordination such as medical training for surgery, learning physical therapy exercises, recreational activities (e.g., martial arts, dance, yoga), and manual skills (e.g., repair, combat training). Recent advances in computer graphics, computer vision, motion capture, and computer power have made it possible to build systems that allow us to assess the effectiveness of fully immersive virtual reality [4, 5].

Full immersion can be achieved though model-based techniques (e.g., capturing a few degrees of freedom and reproducing the captured human motion in a 3D avatar) or image-based techniques (e.g., creating a model of the human from large scale camera arrays and computer vision). Image based methods provide higher degrees of freedom and more accurate representations of the individual at the expense of model quality. Since empirical data supports the notion that motion fidelity is more important and visual fidelity [6], we believe an imaged based methods are most promising for learning fully body motions.

In the current study, we compare learning in an image-based fully immersive VR to instructional videos, the current ubiquitous method of learning full body motion without human interaction. Video-based learning [7, 8] is a good example of mediated learning since it affords instruction on-demand. Videos have a particular advantage over books in that they allow the user to view live, fluid motion of an expert performing a motion. By aggregating the results from 63 separate papers, McNeal and Nelson [9] show that across many different contexts video is a more effective form of instruction than books.

Immersive virtual reality extends the affordances of video, allowing the user to enter the same world as the teacher. First, immersive settings allow users to see in full three dimensions, greatly increasing detail, presence (i.e., learners feel psychologically as if they are in the digital learning environment, as opposed to the physical space [10, 11]), and social presence (i.e., they feel as if the digital reconstruction of the instructor is a real person [12]). Second, as opposed to stationary video, immersive virtual settings allow users to control how they view the environment by allowing them to change aspects such as camera position and orientation.

Third, video settings only allow users to watch the instructor; immersive virtual reality allows the user to interact with the instructor and the environment, as well as to perform novel functions such as sharing body space with the instructor.

In the current work, we compare image-based immersive technology and the established video training tools in their effectiveness in teaching tai chi. We choose tai chi as the learning context because it involves complicated full-body motion and provides clear guidelines for correct performance, and previous work on model-based fully immersive virtual reality has utilized similar learning content [4].

2. Related Work

Since the advent of personal computing and the World Wide Web, there has been a surge of interest in the field of technology mediated learning [13, 14, 15, 16]. Virtual reality has proven to be a promising field for mediated learning, leading to a variety of solutions that address the need for flexible, on-demand training [1]. Virtual training has been met with wide success in the fields of aviation (e.g., flight and space simulators [17]), military (e.g., mission training [18]), medicine (e.g., invasive surgery [19, 20]), emergency (e.g., fire fighting and paramedics [21, 22]), art (e.g., calligraphy [23]), and classroom education [48],

Most empirical research measuring presence between people and virtual humans (either human controlled avatars or computer controlled agents) has utilized systems that track and render the virtual humans along a finite set of degrees of freedom (see [24], for a recent review). In the current work, we are one of the first studies to examine social interaction and social presence using a system in which a user's entire head and body are tracked with extremely high resolution. Our results demonstrate that the high levels of immersion afforded by the fully tele-immersive system result in a higher degree of self-reported presence than video systems.

However, it must be noted the definitions of presence vary between author, and the current research was in no way designed to attempt to provide insights into the intricacies of the semantic definition of presence. Instead, the research took a simulation that was extremely immersive and tested to see if people learned better in it than in a traditional video.

For learning physical motions involving the entire body, it is essential that the user's actions in the physical space correspond to actions rendered on the digital representation of the user (i.e., the avatar) in the virtual space. Yang gives a high level overview of both the hardware and software used in many fully immersive systems, including the image-based system utilized in this study [5], and a typical method of extracting three dimensional models of humans in real-time using off the shelf components is described by Daniilidis and colleagues [25].

In another study, Chua has developed a model-based fully immersive virtual reality system that uses a head mounted display (HMD) along with motion capture to help a student learn tai chi [4]. Their study utilized algorithms that measured deviation between joint movements between the student and the teacher while the student was trying to mimic the teacher, and did not show any advantage in performance for any virtual reality conditions over control conditions. Though Chua and colleagues' experiment provided an important foundation to the study of immersive learning, the current work provides an approach that combines immersive technology with rigorous social scientific methodologies. Consequently, while the previous work did not demonstrate improvements in learning, the current work provides measurement tools sensitive enough to demonstrate the superiority of virtual reality systems over other forms of learning media. There are four notable advancements of the current study over previous work studying learning in virtual reality.

First, because our system is a projection-based system, there is no cumbersome HMD to inhibit naturalistic body motions. We provide extremely high presence without having to drastically limit the user. Second, the image-based reconstruction is not limited to a set number of measured degrees of freedom in the way that model-based motion capture systems are, so the user's physical motions are highly realistic. Third, given that learning such a high level system of movements is likely a gestalt phenomenon, we evaluated our system both via self-report and blind coder ratings. These methods give us the power to test how well participants learn the overall gestalt of the tai chi system of movements, as opposed to only analyzing the success of the learning on a micro, joint-by-joint basis.

Finally, in the current study, we not only measure performance while the virtual teacher is present, but also measure learning, that is, how well the student can perform the tai chi moves later on outside of the simulation when the teacher is no longer present. In sum, the current research is unique both technologically in terms of graphics and methodologically in terms of social scientific learning measurement.

3. Study

We ran a study in our immersive virtual reality system that reconstructed a three-dimensional model of participants as they moved. We placed the participant in a virtual setting where they could interact with their surroundings using natural movement in a tai chi learning environment. Our choice of tai chi was motivated by its slow, fluid motions, which allow us to capture fluid motion in real-time and allow the student to better follow the teacher. Moreover, since tai chi is well established, there is a correct way to perform each move. This allowed us to accurately judge the performance of each participant. Previous research examining virtual performance has utilized similar tasks [4].

3.1. Design

We compared the differences between immersive virtual reality and video as a way to isolate the effect of immersion. Since training videos for physical activities are ubiquitous, we find that this is an apt and crucial comparison for gauging the impact of immersive environments for training physical motion. In this paper the video condition refers to learning from two-dimensional images, and the virtual reality condition refers to interaction in three dimensions with feedback.

3.2. Apparatus

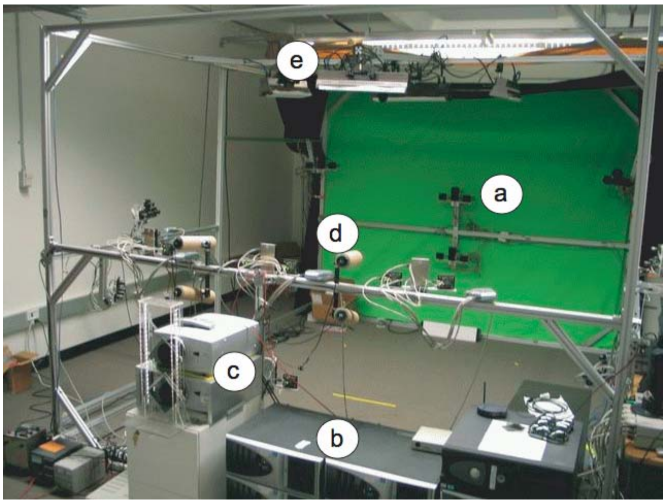

Figure 11: The physical workspace for our immersive reality setting. The camera clusters (a) capture images that are processed by the computers (b) and projected from the projectors (c). The infrared cameras (d) and ceiling lights (e) assist our vision algorithm in capturing a three-dimensional model of the participant.

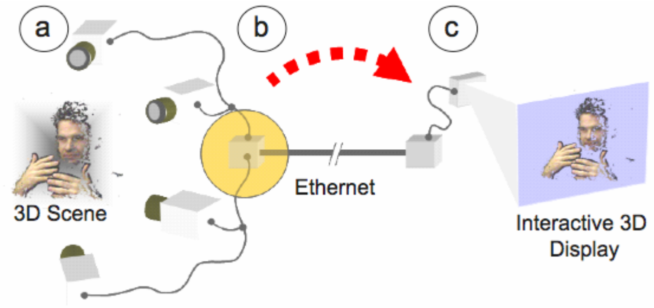

Figure 12: Our system consists of three components image processing (a), data transmission (b), and visualization (c).

Previous work describes our hardware in detail [5]. Figure 1 shows the physical environment in which the participant's image is captured. Our system consisted of three components: image processing, data transmission, and visualization. Figure 2 shows an overview of our system.

In the image-processing phase (Figure 2a), twelve camera clusters (Figure 1a), consisting of three black and white cameras and a color camera, sent images to dedicated computers. Each computer then processed its image (Figure 1b) and computed a 3D representation of the scene from different viewpoints. Partial 3D representation from each viewpoint was combined later in the rendering machine to produce a full representation of the scene. The multiple cameras provided full coverage within a physical workspace of 1.2 meters by 1.2 meters by 2.1 meters.

To enhance the performance of the 3D reconstruction algorithm, ceiling and ground lights illuminated the physical setting (Figure 1e). Infrared cameras (Figure 1d) projected patterns onto objects within the workspace. These patterns did not appear on the model; however the computer used them to increase reconstruction accuracy.

The second component of the system was the network transmission of 3D data (Figure 1b). To achieve immersive virtual environments that allow realistic interactions, 3D streams must be transmitted to a rendering system synchronously with minimal delay.

Figure 13: The projected image of the virtual teacher in our environment. The move name is displayed at the top center of the screen; the reconstructed image is in the middle of the scene.

A simple compression and encoding algorithm reduced bandwidth of data increasing the speed. In visualization phase (Figure 2c), the rendering computer combined the 3D streams from the camera clusters into a single 3D model and placed the model into a virtual environment (Figure 3). Dual projectors (Figure 1c) displayed the virtual environment on a screen in front of the physical workspace. A special pair of polarized glasses combined the overlapping images into one 3D image. Blocking a projector removed one image, which in turn removed depth cues that created the perception of three dimensions. By removing depth, we created the 2D setting for the video condition. Polarized glasses could be worn in either condition without affecting the clarity of image on the screen.

The rendering machine could also record interactions and replay them offline, allowing users to review their performance. A stereo monitor connected to a separate desktop replayed the recorded scene. The monitor simulated two overlapping images by rapidly switching between two images from different viewpoints. As with the projected scene, a special pair of glasses was worn to achieve 3D images of the replay. Image switching was turned off to achieve two dimensions for the video condition.

3.3. Experimental Materials

To create our training program we first captured a 3D model of a tai chi expert as she performed a series of moves. Moves were carefully chosen taking into consideration difficulty and the constraints of the physical workspace and the amount of training time available in a one-hour session. Moves that involved traveling more than a meter or 360-degree rotations were ruled out to accommodate the space restrictions of the workspace while the participant viewed the screen.

The recorded image of the virtual teacher contained three moves, each performed three times with a pause between the discrete moves. Moves were named as follows: Part the Wild Horses Mane, Brush Knee Twist, and Throwing the Loom. While the moves had similar lower body movements, they were quite distinguishable by the upper body movement, especially hand and arm motions.

In both conditions, the virtual world displayed the name of the current move performed by the teacher centered on the top of the screen. In the video condition, the student only saw 2D images of the environment and of the virtual teacher, and only could see the teacher from a single, front-on camera angle (Figure 4). In the virtual reality condition the student saw four stereoscopic human representations, an image of herself rendered in the third person and the teacher from behind, as well as a reflection of both those images in a virtual mirror (Figure 5). We chose a mirror as opposed to arbitrary figures duplicated in space to make the interaction as natural as possible for the participants. Furthermore, the ability to duplicate images of from different vantages is an inherent affordance of our system, and the inclusion of a mirror has been shown to be beneficial when learning physical motion [26].

In order to replay and compare the student's motions during training, the software recorded a model of the student while they were performing the moves. While rendering the scene for review, the virtual mirror was removed, and the student model was synchronized

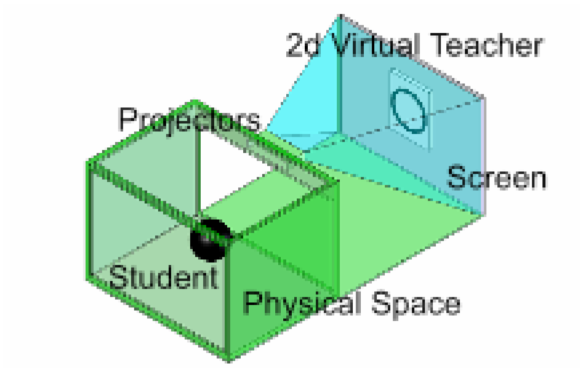

Figure 14: This shows the physical environment and virtual environment in the video condition from a three quarters perspective view. Participants in the physical environment can see a 2D image of the teacher in the screen in front of them.

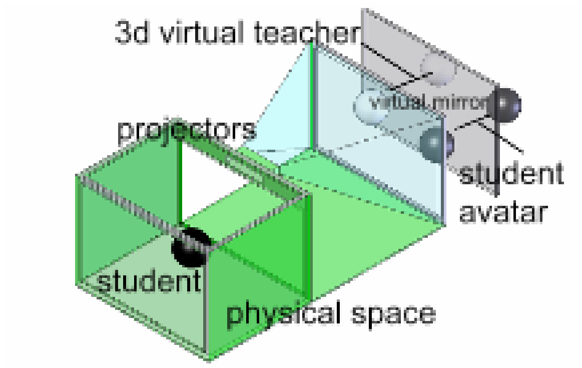

Figure 15: This shows the physical environment and virtual environment in the VR condition from a three quarters perspective view. Students in the physical environment can see the following 3D avatars: the student, the student in the mirror, the teacher, the teacher in the mirror

3.4. Participants

Twenty-six undergraduate students participated in the study and were compensated for their time. The participant pool was split evenly by condition (13 video and 13 virtual reality) and gender (13 male and 13 female). The virtual reality condition contained six male and seven female participants, while the video condition contained seven male and six female participants.

3.5. Procedures

Before entering the physical environment shown in Figure 1, participants were asked to fill out consent from and a demographic survey. After filling out these forms, participants entered the room and were given a briefing of all five phases of the task. Before they began the task, all participants were required to put on the same special clothing that assisted our vision algorithm.

During the first phase, the participants learned tai chi from the prerecorded teacher model. Participants watched and mimicked the teacher to the best of their ability. The environment varied by condition as described in the software section above. In addition to displaying the move names on the screen, the experimenter also called out the name of the move at the beginning of each trial. This phase was videotaped in order to be judged by blind coders during data analysis. In phase two the participant reviewed their performance in phase one on a separate desktop with a stereo monitor. This phase lasted for ten minutes. In the video condition, participants reviewed the actions of the video recording of the teacher in during phase one. The video recording played at the same rate as in phase one, and participants were not able to control any aspect of the recording or playback. In the virtual reality condition, participants received depth cues, control over the recording (e.g., angle, speed, distance), and the ability to examine a three-dimensional rendering of themselves as they interacted with the virtual teacher.

Phase three repeated the same learning task as in phase one. The participant learned from the teacher by watching and mimicking their motions. However, they had the benefit of added learning reinforcement from phase two [27]. Actions in this phase were also videotaped.

Phase four was the same for both conditions. The participants were tested on the individual moves and were asked to recreate the motions without the benefit of seeing the teacher. The experimenter verbally provided the participant with the name of all three moves and the participant performed the moves, one at a time, to the best of his or her ability. This phase was video recorded.

Phase five involved the participant filling out a questionnaire that contained questions about social presence and the usability of the environment. The questions from the questionnaire are listed in Appendix I.

4. Results

4.1. Blind Coder Ratings

Videos recorded during the experiment were separated by participant, phase, and move. For each of the 26 participants, there were three recorded phases with three moves each for a total of nine videos per participant or 234 possible videos in total. Of the 234 total videos, there were four missing or corrupted videos and 11 videos where the participant chose not to perform the task.

Two independent reviewers were trained to judge the specific Tai chi moves, and each evaluator inspected all remaining 219 videos and rated the participants performance. Each video was rated according to 13 separate categories that are depicted in Appendix II.

Of the 13 categories, eight described the steps of the Tai chi move and two described overall form; these were rated on a seven point Likert scale. Coders also rated the participant's knowledge of tai chi and overall performance on a five point Likert scale, and the participant's coordination as either somewhat coordinated or very coordinated. The coder inter-reliability (i.e., how similarly the two coders performed across the various videos and categories) was low but acceptable, Cronbach's α = .55. We then averaged the ratings from the two coders such that, for a given video we had only 13 scores that indicated the mean of the two coders. We then averaged the 13 scores into a single learning measure for each video (Cronbach's α = .98). Table 1 indicates the scores by condition and phase. Participants in the virtual condition consistently outperformed participants in the video condition, especially during the crucial trial of testing which was our strongest measure of actual learning. We also ran analyses to partial out the variance due to individual differences such as age, body size, gender, familiarity with technology, and previous tai chi experience. The difference between learning conditions persists when accounting for these other variables.

	Condition		Significance	
	Video	VR	*t*	*p*
Training I (Phase I)	2.10(0.69)	2.83(1.04)	2.01	0.03
Training II (Phase III)	2.23(0.71)	2.88(1.08)	1.81	0.04
Testing (Phase IV)	2.34(0.54)	3.04(1.18)	1.94	0.03

Table 1: The table above gives the mean and standard deviation values for participants ratings in each condition across all tests. The standard deviation is the value in parenthesis. The **p** ***value (one tailed due to our a-priori, directional prediction) and*** **t** ***scores show students learned better in the VR condition.***

4.2. Self Report

We examined the subjective response questions depicted in Appendix I. We first examined all of the questions surrounding social presence of the instructor. There were ten total questions gauging social presence; we ran a factor analysis using a promax rotation to arrive at a more parsimonious representation of the items. One single factor accounted for 42% of the variance [28], and three questions loaded above .60 on this factor: To what extend did you feel in the same place as the instructor, Did you experience this task as something you did together/jointly with the instructor, and how much the instructor was rated as an expert. We took a mean score of these three questions and ran t-tests to assess the differences between video and VR on social presence ratings. As Table 2 demonstrates, participants reported significantly higher social presence in the virtual reality condition than in the video condition.

We next examined the questions that related to participants analysis of the task itself. There were seven total questions gauging subjective task performance; we ran a factor analysis using a promax rotation to isolate the responses that fell together in a single scale. One single factor accounted for 44% of the variance, and four questions loaded above .60 on this factor: Did you find this task pleasant, How easy was this task, How easy was it to move around in the environment, and ratings of how pleasant the task was. We took a mean score of these four questions, and as Table 2 demonstrates, this difference did not approach significance, and there was no significant effect of gender and no interaction between gender and learning condition. We attributed the lack of significance to the inherent bias that is shown when people evaluate their own behavior [6, 29].

	Condition		Significance	
	Video	VR	*t*	*p*
Social Presence	0.81(0.76)	1.47(1.04)	1.87	0.04
Task Presence	1.81(0.710	1.67(0.58)	0.54	0.30

Table 2: The table above shows the results of our self report questionnaire. The mean and standard deviation for each condition are given with the standard deviation in parenthesis. Ratings of task performance from participants within the virtual setting did not differ significantly from participants within the video setting. However, participants in the virtual setting felt significantly higher social presence.

Conclusions and Future Work

We presented a study assessing the affordances given by immersive virtual environments for technology mediated learning of full body physical motion. Immersive virtual environments were compared to video, the current, ubiquitous method for mediated learning. Through our comparisons, we showed that immersive virtual environments were rated better by blind coders and reported to be better by participants than their video counterparts. We showed that participants within the virtual settings performed better during every phase of the experiment. Self-reports showed participants felt higher social presence within the virtual setting.

In the current study, we designed our control condition of video only to provide a benchmark to compare our immersive system. In doing so, we may have stacked the deck in favor of demonstrating benefits of our immersive virtual reality system. There were learning features in the virtual reality condition (such as adding the mirror in the learning phases and being able to change the angle and distance in the reviewing phases) that were designed to give learners as much advantage as possible. However, these extra features are directly related to the differences between video based systems that we discussed in the introduction. Moreover, as previous work did not demonstrate reliable differences with different types of virtual tai chi instruction systems [4], we wanted to first demonstrate significant differences and then, in future work, scale the differences between virtual reality and video back in order to isolate the theoretical variables that contributed to the improved learning (e.g., stereoscopic viewing versus viewpoint control versus rendering the self in third person).

Our system captures three-dimensional models of motion in real-time, however the quality and speed of reconstruction is far from perfect. Increasing the number of cameras and infrared projectors increases coverage and redundancy leading to denser, more accurate images. Cameras that identify, focus on, and follow certain body parts could increase the accuracy in key regions (e.g., face, hands) [30, 31].

The speed of our current system is constrained by two main factors. First, extracting and combining partial models from cameras clusters is computationally expensive. Porting vision algorithms to existing graphics hardware could improve speed in this regard [32, 33]. Second, the sheer volume of data saturates the network and storage bandwidths. Using techniques that have low bandwidth consumption help ease the data management load [34, 35].

The current system reconstructs both the teacher and student models using the same real-time algorithm, however only the student model needs to be captured in real-time. Real-time performance comes at the expense of reconstruction quality; by using offline algorithms to overlay meshes on the teacher model, we can fill in holes and increase the accuracy making it easier for the students to distinguish subtle differences in the teacher's motion [36, 37, 38].

During the learning phases, participants watched and mimicked the virtual teacher. This level of interaction is acceptable for tai chi; however, other full body physical activities require feedback and the ability to manipulate the environment. Interactivity can be achieved by detecting collisions between manipulators (e.g., hands, tools) and virtual objects [39, 40, 41]. Small vibration units can give haptic feedback, indicating successful interaction with a virtual object [42, 43]. For example, a participant could dance with a virtual partner and know that their avatars are making contact; or a participant could swing a golf club to hit a virtual ball. Increased interactivity can also be achieved by increasing the intelligence of the virtual agent. The current system uses a pre-recorded model as the teacher. While human level artificial intelligence is far beyond our capabilities, it can be approximated by creating simple reactive avatars or simulated by remotely controlling the teacher model. It has been shown that the simple actions such as mimicking non-verbal gestures are enough to give the perception of intelligence [44]. As such, encoding these actions within the teacher model can increase the feeling of immersion. A simpler approach involves connecting a remote teacher to student by networking two immersive reality systems [5]. The remote teacher can interact though natural motion without having to be collocated with the student. Comparative studies can assess the benefits of each approach.

Exciting work has been done regarding collaboration within virtual environments [45, 46], however space constraints of the current physical environment limit both the number of participants in the virtual environment and the amount any given participant can move. Since each participant retains their full mobility in the virtual world, enabling our physical environment to support multiple people would expand our learning domain to include collaborative tasks (e.g., disaster relief, construction, and surgery). Retaining natural movement bolsters the validity of statements about virtual environments and the effect on social interaction and learning.

In the current system, the participant must look at a screen in front of them to see the virtual world. Exploring other immersive technologies such as HMDs or full CAVE [47] environments may increase mobility and lead to a higher degree of immersion and better learning. Different technologies are better suited to different tasks [48], studying the interaction between technology and task can help build better systems.

Changes in technology can be assessed to discover their effects on learning physical motion. As interactivity and realism of our environment increases, we hope to bridge the gap between learning from a virtual teacher, in a virtual environment, to learning from face-to-face interaction with a real teacher. In the current work, we have demonstrated persuasive evidence that immersive virtual reality provides better learning of physical movements than a two-dimensional video. As technology, and our understanding of how to use that technology, improves we should see larger gains in learning from virtual reality.

Acknowledgements

We would like to thank the Stanford LIFE lab and Nick Yee for helpful suggestions on the design of this research. We thank Nancy Ru for providing us with our teacher model, Alice Kim, Zi Han, and Claire Carlson for encoding and inputting data, and Jaireh Teccaro for creating figures.

References

[1] F Mantovani, *Virtual Reality Learning: Potential and Challenges for the Use of 3D Environments in Education and Training*, chapter 12, Towards Cyber-Psychology: Mind, Cognitions and Society in the Internet Age. IOS Press, Amsterdam, 2001.

[2] J Bruner, *Towards a theory of instruction*, WW Norton, New York, 1966.

[3] E Trondsen and K Vickery, "Learning on demand," *Journal of Knowledge Management*, vol. 1, no. 3, pp. 169–180, March 1997.

[4] P T Chua, R Crivella, B Daly, N Hu, R Schaaf, D Ventura, T Camill, J Hodgins, and R Pausch, "Training for physical tasks in virtual reality environments: Tai chi," in *Proc of IEEE Virtual Reality*. IEEE, 2003, pp. 87–94.

[5] Z Yang, K Nahrstedt, Y Cui, B Yu, J Liang, Sang Hack Jung, and Ruzena Bajscy, "Teeve: The next generation architecture for tele-immersive environments," in *Proc IEEE International Symposium on Multimedia*, Irvine, CA, USA, 2005, IEEE.

[6] J N Bailenson, K R Swinth, C L Hoyt, S Persky, A Dimov, and J Blascovich, "The independent and interactive effects of embodied agent appearance and behaviour on self-report, cognitive, and behavioural markers of copresence in immersive virtual environments.," *PRESENCE: Teleoperators and Virtual Environments*, vol. 14, pp. 379–393, 2005.

[7] R D Pea, Video-as-data and digital video manipulation techniques for transforming learning sciences research, education and other cultural practices, *International Handbook of Virtual Learning Environments*. Kluwer Academic Publishing, 2005.

[8] Daniel L Schwartz and K Hartman*, It is not television anymore: Designing digital video for learning and assessment*, Video research in the learning sciences. Erlbaum, Mahwah, NJ, 2006.

[9] B J McNeal and K R Nelson, "Meta-analysis of interactive video instruction: A 10 year review of achievement effects," *J. COMP. BASED INSTR*, vol. 18, no. 1, pp. 1–6, 1991.

[10] J M Loomis, "Distal attribution and presence," *PRESENCE: Teleoperators and Virtual Environments*, no. 1, pp. 113–119, 1992.

[11] M Slater, "Presence and the sixth sense," *PRESENCE: Teleoperators and Virtual Environments*, vol. 11, no. 4, pp. 435–439, 2002.

[12] J Blascovich, J Loomis, A Beall, K Swinth, C Hoyt, and J Bailenson, "Immersive virtual environment technology as a methodological tool for social psychology," *Psychological Inquiry*, vol. 13, pp. 103–124, 2002.

[13] M Alavi, B C Wheeler J S, and Valacich, "Using it to reengineer business education: An exploratory investigation of collaborative telelearning," *MIS Quarterly*, vol. 19, no. 3, pp. 293–313, 1995.

[14] G Piccoli, R Ahmad, and B Ives, "Web-based virtual learning environments: A research framework and a preliminary assessment of effectiveness in basic it skills training," MIS *Quarterly*, vol. 25, no. 4, pp. 401–426, 2001.

[15] M-N Lamy and R Goodfellow, "Reflective conversation," *Virtual Language Classroom. Language Learning and Technology*, vol. 2, no. 2, pp. 43–61, 1999.

[16] M Warschauer, *Interaction, Negotiation and Computer-mediated Learning*, Educational Technology in Language Learning: Theoretical Considerations and Practical Applications. 1998.

[17] R Lofin and P Kenny, "Training the hubble space telescope flight team," *IEEE Computer Graphics and Applications*, vol. 15, pp. 31–37, 1995.

[18] Albert Rizzo, Jacquelyn Ford Morie, Josh Williams, Jarrell Pair, and J Galen Buckwalter, "Human emotional state and its relevance to military VR training," in *International Conference on Human Computer Interaction*, 2004.

[19] J Berkley, G Turkiyyah, D Berg, M Ganter, and S Weghorst, "Real-time finite element modelling for surgery simulation: An application to virtual suturing," *IEEE Transactions on Visualization and Computer Graphics*, vol. 10, no. 3, pp. 1–12, 2004.

[20] S Delp, P Loan, C Basdogan, and J Rosen, "Surgical simulation: An emerging technology for training in emergency medicine," *PRESENCE: Teleoperators and Virtual Environments*, vol. 6, pp. 147–159, 1997.

[21] S Stansfield, D Shawver, and A Sobel, "Medisim: A prototype virtual reality system for training medical first responders," in *Proc of IEEE Virtual Reality*. IEEE, 1998.

[22] D Tate, L Silbert, and T King, "Virtual environments for shipboard firefighting training," in *Proc IEEE Virtual Reality Conference*. IEEE, 1997.

[23] U Yang, "Just follow me: An immersive virtual reality based motion training system," in *International Conference on Virtual Systems and Multimedia*, 1999.

[24] K M LEE, "Presence, explicated," *Communication Theory*, vol. 14, pp. 27–50, 2004.

[25] K Daniilidis, J Mulligan, R McKendall, G Kamberova, D Schmid, and R Bajcsy, *Real-Time 3D Tele-immersion*, pp. 253–266, The Confluence of Vision and Graphics. Kluwer Academic, 2000.

[26] L P Sewall, T G Reeve, and R A Day, "Effect of concurrent visual feedback on acquisition of a weightlifting skill," *Percept Mot Skills*, vol. 67, no. 3, pp. 715–718, December 1988.

[27] D L Weeks and L P Anderson, "The interaction of observational learning with overt practice: effects on motor skill learning," *Acta Psychol*, vol. 104, no. 2, pp. 259–271, May 2000.

[28] A K Romney, S Weller, and W Batchelder, "Culture as consensus: a theory of culture and informant accuracy," *American Anthropologist*, vol. 88, pp. 313–338, 1986.

[29] [29] R E Nisbett and T D Wilson, "Telling more than we can know: Verbal reports on mental processes," *Psychological Review,* vol. 84, pp. 231–259, 1977.

[30] Henry Schneiderman, "Learning a restricted bayesian network for object detection," in *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, June 2004.

[31] Ying Wu, Gang Hua, and Ting Yu, "Tracking articulated body by dynamic markov network," in *IEEE International Conference on Computer Vision*, Nice, France, October 2003, IEEE, pp. 1094–1101, IEEE Computer Society.

[32] James Fung and Steve Mann, "Openvidia: parallel gpu computer vision," in *MULTIMEDIA '05: Proceedings of the 13th annual ACM international conference on Multimedia,* New York, NY, USA, 2005, pp. 849–852, ACM Press.

[33] William R. Mark, R. Steven Glanville, Kurt Akeley, and Mark J. Kilgard, "Cg: a system for programming graphics hardware in a c-like language," *ACM Transactions on Graphics (SIGGRAPH 2003)*, vol. 22, pp. 896–907, 2003.

[34] Marc Levoy and Pat Hanrahan, "Light field rendering," in *SIGGRAPH '96: Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, New York, NY, USA, 1996, pp. 31–42, ACM Press.

[35] Steven J. Gortler, Radek Grzeszczuk, Richard Szeliski, and Michael F. Cohen, "The lumigraph," in *SIGGRAPH*

'96: Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, New York, NY, USA, 1996, pp. 43–54, ACM Press.

[36] D. Anguelov, P. Srinivasan, D. Koller, S. Thrun, J. Rodgers, and J. Davis, "SCAPE: Shape completion and animation of people," in *Proceedings of the 32nd International Conference on Computer Graphics and Interactive Techniques (SIGGRAPH),* Los Angeles, California, August 2005.

[37] Diego Nehab, Szymon Rusinkiewicz, James Davis, and Ravi Ramamoorthi, "Efficiently combining positions and normals for precise 3d geometry," *ACM Transactions on Graphics (SIGGRAPH 2005)*, vol. 24, no. 3, Aug 2005.

[38] Greg Turk and Marc Levoy, "Zippered polygon meshes from range images," in *SIGGRAPH '94: Proceedings of the 21st annual conference on Computer graphics and interactive techniques*, New York, NY, USA, 1994, pp. 311–318, ACM Press.

[39] Ming C. Lin, "Fast and accurate collision detection for virtual environments," in *Scientific Visualization*, Washington, DC, USA, 1999, IEEE, pp. 171–180.

[40] Naga K. Govindaraju, David Knott, Nitin Jain, Ilknur Kabul, Rasmus Tamstorf, Russell Gayle, Ming C. Lin, and Dinesh Manocha, "Interactive collision detection between deformable models using chromatic decomposition," *ACM Transactions on Graphics (SIGGRAPH 2005)*, vol. 24, no. 3, pp. 991–999, 2005.

[41] Jan Klein and Gabriel Zachmann, "Point cloud collision detection," in *Computer Graphics forum (Proc. EUROGRAPHICS)*, M.P. Cani and M. Slater, Eds., Grenoble, France, Aug 30 – Sep 3 2004, number 3 in 23, pp. 567–576.

[42] Karon E. MacLean, "Designing with haptic feedback," in *Proc of International Conference on Robotics and Automation*, San Francisco, CA, USA, April 2000, IEEE.

[43] Allison M. Okamura, Stanford Dennerlein, and Robert D. Howe, "Vibration feedback models for virtual environments," in *Proc of International Conference on Robotics and Automation*. IEEE, 1998, pp. 674–679.

[44] Jeremy N Bailenson and Nick Yee, "Digital chameleons: Automatic assimilation of nonverbal gestures in immersive virtual environments," *Psychological Science*, vol. 16, pp. 814–819, 2005.

[45] Veronique Normand, Christian Babski, Steve Benford, Adrian Bullock, Stephane Carion, Yiorgos Chrysanthou, Nicolas Farcet, John Harvey, Nico Kuijpers, Nadia Magnenat-Thalmann, Soraia Raupp Musse, Tom Rodden, Mel Slater, and Gareth Smith, "The coven project: Exploring applicative, technical, and usage dimensions of collaborative virtual environments," *PRESENCE: Teleoperators and Virtual Environments*, vol. 8, no. 2, pp. 218–236, 1999.

[46] Jaron Lanier, "Virtually there," *Scientific American*, pp. 66–75, April 2001.

[47] C Cruz-Neira, D J Sandin, and T A DeFanti, "Surround-screen projection-based virtual reality: The design and implementation of the cave," in *Proc of SIGGRAPH-93: Computer Graphics*, Anaheim, CA, 1993, pp. 135–142.

[48] Pantelidis, Veronica S. Virtual Reality in the Classroom. *Educational Technology* vol. 33 no. 4 (April 1993), p. 23-27. (General.) Reprinted in John Hirschbuhl, ed., *Computer Studies: Computers in Education*, p.175-179. Guilford, Connecticut: The Dushkin Publishing Group, Inc., 1993.

Appendix

I. Self Report Questions

The following questions were rated on a five point Likhert scale. The extremes of each scale are in brackets.

- To what extent did you have the sense that you were in the same place as the instructor? [from a very small extent to a very large extent]
- How would you rate your awareness of the instructor's intentions/wishes in this task? [from low awareness to high awareness]
- Did you find this task pleasant or unpleasant? [from very pleasant to very unpleasant]
- Did you experience this task as something that you did together/jointly with the instructor, or as something you did on your own/separately)? [from a very large extent on my own to every large extent together]
- How easy or difficult was this task? [from very difficult to very easy]
- How easy or difficult was it to move around in the environment? [from difficult to very easy]The following questions were rated on a five point Likhert scale. The extremes of each scale are in brackets.
- To what extent did you have the sense that you were in the same place as the instructor? [from a very small extent to a very large extent]
- How would you rate your awareness of the instructor's intentions/wishes in this task? [from low awareness to high awareness]
- Did you find this task pleasant or unpleasant? [from very pleasant to very unpleasant]
- Did you experience this task as something that you did together/jointly with the instructor, or as something you did on your own/separately)? [from a very large extent on my own to every large extent together]
- How easy or difficult was this task? [from very difficult to very easy]
- How easy or difficult was it to move around in the environment? [from difficult to very easy]
- For each of the word pairs below, please circle the number that best suits your experience of the learning environment. [from personal to impersonal, from social to antisocial, from lively to lifeless, and from pleasant to unpleasant]
- For each of the word pairs below, please circle the number that best suits your experience of the instructor during the task that you have just done. [from close to distant, from responsive to unresponsive, from active to passive, from warm to cold, from helpful to unhelpful, from realistic to fake, and from an expert to a novice]

II. Blind Coder Rating Criteria

The following questions were rated on a seven point Likhert scale from very poor to very well. Coders were given guidelines on how to evaluate each step for each move and how to evaluate fluidity and posture.

- You are provided with descriptions of how to perform each move broken down into distinct steps. On the scale below rate the participant's performance on each step of the move.
- To correctly perform tai chi each move part must flow smoothly into the next. In addition, the participant must have good posture. Good technique is described in the attached handout. On the scale below rate the fluidity of the moves and the participant's posture throughout the entire move.

The following questions were rated on a five point Likhert scale from poor to excellent.

- Please rate the participant's knowledge of tai chi
- Please rate the participant's overall tai chi performance

For the final question the coders rated the participants overall coordination as either somewhat coordinated or very coordinated.