Onebody: Remote Posture Guidance System using First Person View in Virtual Environment

Thuong N Hoang **Martin Reinoso** **Frank Vetere** **Egemen Tanin**

Microsoft Research Centre for Social Natural User Interfaces
Department of Computing and Information Systems
The University of Melbourne, Australia
{thuong.hoang, martin.reinoso, f.vetere, etanin}@unimelb.edu.au

ABSTRACT

We present Onebody, a virtual reality system for remote posture guidance during sports or physical activity training, such as martial arts, yoga or dance, using first person perspective. The system uses skeletal tracking of the instructor and the students, rendered as virtual avatars. Using a virtual reality headset, the student can visualise the movement of the instructor's avatar, rendered in place of their own body. Onebody provides a first person perspective of the movement instruction, allowing the student to *step into the instructor's body*. We conducted a study to compare the performance of Onebody in terms of posture matching accuracy and user's preference, with existing techniques of delivering movement instructions, including pre-recorded video, video conferencing and third person view virtual reality. The result indicated that Onebody offers better posture accuracy in delivering movement instructions.

Author Keywords

Virtual reality; posture guidance; first person view.

ACM Classification Keywords

H.5.1 [Multimedia Information Systems]: Artificial, augmented, and virtual realities; H.5.2 [User Interfaces]: Evaluation/Methodology

INTRODUCTION

The process of motor skill learning is observed in multiple training domains, such as in sports, martial arts, or dance. Researchers have identified multiple factors affecting the performance and learning of motor skill, such as observation practice, focus of attention, and feedback [17]. During motor training, one of the main roles of the instructor is to provide feedback to the learner. Such information comes from a source external to the learner and cannot be extracted from the learning environment, thus is called *augmented* or *extrinsic feedback* [13]. The feedback can be provided either as the action is being performed, called *concurrent feedback*, or after the action has been completed, called *terminal feedback*. Concurrent feedback has been demonstrated to increase performance in terms of skill acquisition during training for the novice learner [12].

In the context of martial arts or similar forms of posture or movement training, such as yoga or dance, feedback also takes the form of movement guidance through a process of demonstration and mimicry [19]. An instructor performs a particular movement with verbal instructions, and a student or class of students observe and imitate the movements. The instructor provides augmented feedback verbally or via physical interactions with the student. Verbal instruction has limited capability to accurately describe movements or feedback, and can hinder the learning process by overloading information [20]. Concurrent feedback through physical interaction is challenging, if not impossible, as the instructor can only correct the student's movement after demonstration is completed. Further challenges arise when providing augmented feedback in the context of remote motor skill learning, i.e. when the instructor and the student are not collocated. We aim to ameliorate these limitations of concurrent feedback in the physical activity training domain.

Recent advances in display technologies, such as virtual reality (VR), have been utilised to provide training in many motor skill learning contexts, including sports [5, 16], dance and martial arts [2, 3, 11], and rehabilitation training [4, 6, 8, 14]. One of the major advantages of virtual reality training systems is the ability to provide concurrent feedback visually and immersively through head mounted displays (HMD), large screens or projections. Among the display technologies, HMD provides the fullest immersive experience by covering the entire field of vision of the user. Previous works explore the use of HMD to provide a first person perspective view for motor skill training in many application areas [6, 11, 21]. Instructions to perform a particular skill are recorded by an expert, and superimposed directly on the learner's virtual view. The resulting visualisation for the learner is a first person view of the movement as if seen from the eyes of the instructor. We are inspired by this visualisation capability of VR using HMD for motor skill training. We are particularly interested in the

NordiCHI '16, October 23-27, 2016, Gothenburg, Sweden

DOI: http://dx.doi.org/10.1145/2971485.2971521

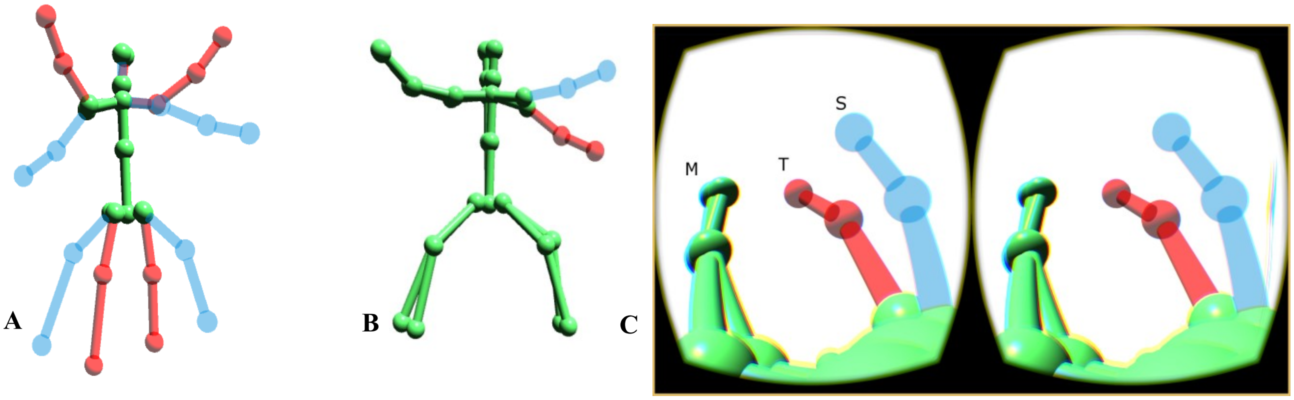

Figure 1. **A and B: Overlapping Avatars; C: Student's view of instructor body through an Oculus Rift: S indicates the student's arm, T teacher's and M is the matched arm**

application for synchronous learning where the instructor and the learner or student interacts in real time, as compared to asynchronous interaction through pre-recorded instructions. Existing research has investigated synchronous motor learning in immersive VR [4, 11]; however, there is limited empirical evidence demonstrating the benefits of synchronous first person view VR for motor training.

In this paper we present a virtual reality system for remote posture guidance providing concurrent feedback through first person perspective, called *Onebody*. The contribution of the paper includes the Onebody posture guidance system and the empirical validation of first person view in terms of posture accuracy. Previous studies [2, 7, 9, 11] did not present conclusive results regarding first person view and virtual reality training. In our study, we explored different aspects of Onebody as a posture guidance system, including synchronous interaction, VR environment, and first person perspective, in a single experiment.

BACKGROUND

Immersive environments have been utilised for providing movement guidance as concurrent feedback in many previous research work. Multiple aspects of the virtual environment are the focus of many previous work, including asynchronous learning, first person view, and the virtual environment. The study presented in this paper encompassed all of those aspects.

Asynchronous interaction

A study was conducted for teaching Tai Chi movement using virtual reality by Chua et al. [11]. Instructors and students are rendered as realistic humanoid avatars in the virtual environment. The movement of the instructor was pre-recorded, supporting only asynchronous learning. The researchers completed a study to compare the effectiveness of different arrangement of the avatars, including surrounding the teacher, side by side with teacher, and superimposed on the student with wireframe and stick figure. All conditions used virtual reality. However, they did not find any significant result for the superimposed condition. Their study discussed the difficulty in measuring error and time delay for comparing motor movements of the teacher and student. Onebody focuses on posture accuracy, using the movement as instructions to achieve the postures.

The work by Bailenson et al. [2] and related previous work [7, 9] compared VR and video training as a training medium for Tai Chi. The VR condition is a mirroring third person view, with reconstructed 3D avatar from 360 images capture. The objective measure used was blind coders, who graded the video recording of participants' performances, as compared against a video of the instructor. The VR system was implemented using projectors, instead of head mounted display. The study showed only significant advantage of VR training through self-reported measures.

Both studies explored asynchronous learning without real time interaction with an instructor. We are interested in exploring synchronous interactions where the instructor and the student both perform the movement simultaneously.

First person view

Virtual reality technology can simulate multiple viewpoints for the user, including first person or third person perspective. Many previous studies explored different aspects of virtual reality physical activity training, including the VR medium and superimposed visualisation or first person view.

A VR training system was developed by Yang and Kim [21] called Just Follow Me. They conducted a study to evaluate the performance of a VR system using head mounted display that renders a ghost avatar from a first person view for motion training. The comparison was made against video-based training method. The result did not indicate a strong advantage of VR over video based method in terms of accuracy.

Ghostman [4] is an augmented reality system that overlays instructional video from a first person view, and is used for remote and telerehabilitation instructions of motor skill. The authors performed a randomised controlled pilot study comparing the Ghostman system with face to face teaching. There was no significant improvement on skill errors nor learning effect between the conditions.

Virtual environment

Virtual and augmented environments provide the visualisation capability that is beneficial to motor skill trainings. There are many examples of previous work in this area, notably in the area of rehabilitation and sport training.

YouMove [1] provides an interactive AR mirror where users can record and learn physical movements. Acting as a mirror, the system provides an external view of pre-recorded movements in the forms of stick figures for the user to match. Kallman, Camporesi, and Han [8] implemented a VR system using a large display to provide instructions for physical rehabilitation programs. A similar system by Tang et al. [14] uses body tracking technology for guidance and feedback for physiotherapy exercise. Their system implements visual guides and multiple camera views to improve exercise accuracy. They used Microsoft Kinect for body tracking, which has previously been demonstrated to be suitable in a clinical setting [15].

Chan et al. [3] implemented a screen-based virtual reality system to teach dance movement, using a retroreflective motion tracking system. The study shows that their system successfully guided students to improve their skills as compared to tradition video learning, for dance movements. An immersive projection system was used. The study shows successful skills improvement and suggest to improve their system by the use of more immersive equipment like a head mounted display.

Previous works also cover skill based training with sports equipment. Covaci et al. [5] performed a comparison of third and first person view for basketball training using immersive virtual environments, using a physical ball. Their implementation of virtual reality is through a large screen image display platform using projectors, instead of head mounted display. Their study did not show any significant different in player's performance compared between the first and third person view. In contrast to Covaci and colleagues' work where physical equipment was used, Todorov et al. [16] built a simulated table tennis game and provided virtual overlay of a reference racket movement. They performed user studies to compare the participant's performance after training in a physical environment and the simulated environment. The results indicated improvement in task performance under the simulated environment with augmented feedback.

Research gap

Onebody emerges in an established research literature regarding motor skill learning, augmented feedback, and virtual reality training systems. The contribution of Onebody with regards to previous work is the validation of motor skill training theory [6, 12, 18] regarding the benefits of concurrent visual feedback to improve error performance for the novice learner in a synchronous learning context. Our work provides empirical evidence of significant performance improvement, which was absent in previous studies [2, 9, 11]. Onebody uses head mounted display for full immersiveness, as compared to projection based VR [5] and other non-immersive systems [3]. Prior research has focused on separate aspects of motor skill learning, including asynchronous [2, 11] and synchronous interaction [4, 21], and first person view [5, 16]. Our work explores different aspects of Onebody as a posture guidance system in a single study, including synchronous interaction, VR environment, and first person perspective. We are interested in the performance aspects of skill training. Skill retention is not the focus of our research.

ONEBODY

Onebody is a virtual reality system for remote posture guidance using first person perspective. The system utilises the Microsoft Kinect sensor for skeletal tracking of an instructor and a student, who are not collocated. By overlaying the virtual avatars of the instructor and the student, the system creates a visualisation of first person perspective to deliver movement instructions.

Visualisation

Movements of both the instructor and student are captured by Kinect sensors and rendered as virtual avatars in a virtual environment. The avatars are superimposed in place and normalised so joint dimensions are matching. Using the Oculus virtual reality headset, the student can view and mimic the movements of the instructor from a first person perspective. As the instructor demonstrates a stance, the student can visualise the instructor's movements as virtual limbs extending from their own body. The student then moves their own virtual body to match and learns to achieve the final posture. In other words, the resulting visualisation can be described as *stepping into the instructor's body*. Thus, the system is called *Onebody*. The instructor utilises the same first person visualisation through a VR headset to provide real time feedback and assist the student in reaching the stance. Onebody provides a medium to deliver body movement instructions for non-collocated instructor and learner. We are particularly interested in application of Onebody for remote training. Remote training for posture or movement is challenging due to the physical separation of the instructor and the student, which results in limitation in delivering body movement instructors as well as providing concurrent feedback.

Infrastructure

Onebody is a posture guidance system, consisting of 2 stations: the instructor and the student station. Each station has exactly the same hardware configuration. The user of each station, either the instructor or the student, will be standing in front of a Microsoft Kinect sensor while wearing an Oculus Rift head mounted display.

Through the Oculus, the instructor and the student are brought into a shared immersive virtual reality environment,

Figure 2. A student participant (left) and the instructor (right) during the study

where each is represented by a stick figure avatar. The movement of each user is mapped directly to the corresponding avatar through skeletal tracking by the Microsoft Kinect sensor.

Within the virtual world, Onebody attaches both avatars at the hip while still maintain the mapping between users and corresponding avatars. The resulting effect is a visualization where each user can observe the movements of the other person overlaid in place. The hip attachment allows the student to observe the instructor's virtual body extending from their own as if seen from the first person view of the instructor. Figure 1C shows the view of the student through the Oculus Rift, with the instructor's arms in red and the overlapping body parts in green.

Posture guidance

We use colour coded visualisation for the avatars to indicate posture matching, as a form of concurrent feedback. The student's avatar is blue, instructor's is red and the parts of the virtual bodies that overlap turns green, as shown in Figure 1A and B. Transparent virtual avatars are used so that both colored avatars will be visible.

We implemented a matching threshold by comparing the distance of the 3D position of joints data between the instructor and the student. If the distance difference between the joints are under the threshold of 5cm, the overlapping avatar will turn green. This threshold was implemented to reduce tracking errors and jittering. Figure 1B shows the majority of the body overlapped within the threshold.

The colour code system provides a form of concurrent visual feedback for the student. The error estimate of the student performance is displayed in real time and communicated simultaneously to both the student and the instructor through the colour of the overlayed avatars. Continuous visual feedback has been demonstrated to improved performance by reducing error during practice in the learning of discrete simple movement tasks [12, 18]. However, training with concurrent feedback has a negative impact on skill retention, especially after the feedback has been removed. The limitation of Onebody as a guidance system is similar to that of continuous visual feedback. The scope of our Onebody system is only aiming at the early stage of skill learning, to deliver instructions in the most effective method. We are not targeting skill retention with Onebody.

Infrastructure and Body normalisation

Onebody implements a normalisation to skeleton tracking to match the difference in body sizes. The instructor can perform any body posture or movement and the algorithm will translate the movements into movements of a virtual body that has the student's body measurements. Normalisation is achieved by using normalised vectors to represent body segment information.

Onebody uses Client-Server as the network architecture to support remote training. Each user, either an instructor or a student, is a client, and a central server communicates the information between all clients. Skeletal tracking information from Kinect sensors of each client is sent directly to the server, who broadcasts to other clients. Comparison of skeletal tracking is performed at client's side to display the colour-coded avatar view for concurrent feedback.

EVALUATION

We designed a user study to evaluate the performance of Onebody in terms of posture accuracy and user's preference.

Figure 3. Sample postures for the study

Methodology

Our goal in this study is to evaluate the performance of Onebody as an instructional medium for posture guidance. Our main hypothesis is:

H1: Onebody delivers better posture accuracy than existing remote movement instruction methods.

Posture accuracy is determined by the extent to which the student can replicate the final posture as instructed and demonstrated by the instructor. We also explored performance in terms of interaction time and subjective instructor's score for student's posture.

In addition to the main hypothesis, we are interested in user's perception and preference, in terms of ease of use, ease of understanding, perceived accuracy and user preference.

Participants

The participants for the study include one instructor and multiple students. The instructor is required to have prior experience in teaching postures. We contacted the Department of Physiotherapy at our university and sent out a call to final year students of the post graduate physiotherapy course, who have successfully completed the posture class as part of their coursework. The screening was completed by the lecturers at the department based on our defined criteria that the candidate must have experience in teaching postures. We selected one suitable candidate to be the instructor for the entire study, to ensure consistency of instructions across all student participants. We met with the instructor and trained him on the methods for 3 hours. The instructor was recruited as a contractor paid by the hour.

We recruited 23 participants (13 males and 10 females, aged from 17 to 32, mean 23, SD 3.8) to play the part of the student in the training sessions. There was no financial incentive for the participants. The participants had previous background in a range of sports, including ballet, martial arts, and yoga, with moderate experience.

Design

To test our hypothesis, we compared Onebody against other remote methods for posture training in terms of performance and user preference. The methods for training is the independent variable, and the dependent variables are the performance factors of posture accuracy, completion time, subjective instructor's rating, and questionnaire for user's preference. The main task for the participants is a remote training session between an instructor and the student participant, to learn a number of martial arts postures or stances.

Independent variable

The independent variable is the different methods for remote posture training. We selected 3 other methods, including pre-recorded video (*PRV*), video conferencing (*Skype*) and virtual reality with 3rd person perspective (*VR-3PP*, see Figure 4). Each of the methods differs to Onebody (*OB*) in a single aspect, namely synchronous interaction, VR medium, and 1st person perspective. By performing pairwise t-test of the methods, we aim to gain insights into which aspects of Onebody contribute to performance gain (if any). The summary of differences is shown in 0.

	Video	*Skype*	*VR-3PP*	*Onebody*
Synchronous Interaction	No	Yes	Yes	Yes
VR Medium	No	No	Yes	Yes
1st Person View	No	No	No	Yes

Table 1 Differences in training methods

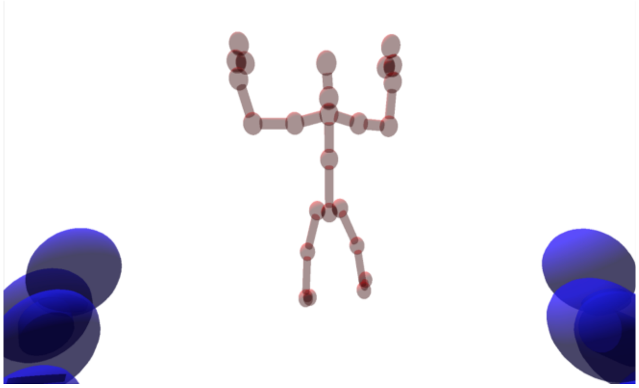

Figure 4. Student's view in the VR-3PP condition

Dependent variables

In order to evaluate the performance of Onebody, we choose the task of a remote training session of martial arts posture. The main outcome of the task is the student's ability to match the posture as demonstrated and guided by an instructor. To evaluate performance, we measure *posture accuracy*, *completion time*, and *subjective score* by the instructor, as well as user preference through a questionnaire.

It is important to note that we were focused only on performance and preference factors. Our study design and duration did not aim to evaluate other factors such as motor skill retention and motor learning.

Task

We designed a 4x4 within subject study. The main activity of the study is a training process between an instructor and the student participant, who are remotely located using the 4 methods in question (video, Skype, VR-3PP, and Onebody).

The student participant took part in one training session, using one of the training method. The instructor, who was not collocated, provided verbal guidance to move different parts of the body, such as "*Extend your left arm 90 degrees in front",* and performed such action at the same time. The student mimicked the movement of the instructor. Through each method, the instructor could observe the student's movement through either a video feed or in a VR environment. The instructor provided verbal feedback if necessary for the student to correct the movement. The student could ask the instructor to repeat the instructions if required. Once the student was confident with the movement, she or he indicated to the researcher located in the same room to capture the final posture. The student and the instructor would move into the final posture and joints position data of both were captured via skeletal tracking by Kinect sensors. A matching accuracy percentage was calculated from this data, as an indication of posture accuracy.

There were 4 sets of postures, each with different complexities, referred to as postures A, B, C, and D. For each training method, the participant completed all four posture types (4x4 design). For each of posture types, there were 4 variants so that the participant did not perform any posture twice, thus reducing any learning effect. The variants maintained the complexity level with changes to the defined angles of the arms, and were called, for example, A1, A2, A3, and A4. The order of the methods and the order of the variants were randomized. In total, each participant completed 16 posture training sessions. Samples of the postures are shown in Figure 3. The postures were inspired by martial art stances.

Data recording

For each posture, the following information was recorded: posture accuracy, completion time, and subjective instructor score.

Posture accuracy was recorded as the extent to which (percentage) the student managed to match the posture of the instructor. Skeletal tracking of the student and the instructor were captured for each posture as previously described. The 3D positions of the following joints were recorded: wrist, elbow, shoulder, hip, knee and ankle. For the skeletal data, each limb was made up of 2 bone segments. Each bone segment was stored as a normalised vector calculated from the 3D position of corresponding joints. A baseline position was considered a rest position where both arms were at rest alongside the body and the legs were standing up straight. For each bone segment, the difference angle between the instructor's and the student's skeletal data was calculated. This difference angle was mapped to a percentage between the baseline position and the instructor's correct position. In other words, if the participant assumed the rest position, the score would be 0. The posture variances within a set were designed so that there were similar angular differences between the correct position and the baseline position. A weighting of 25% was applied to each pair of bone segments that made up each of the four limbs. In other words, the percentage score for each limb ranged from a minimum of 0 to a maximum of 25%. Totalling the score of 4 limbs would create a maximum score of 100% for posture accuracy.

An *instructor score* from 1 to 10 was manually recorded by the instructor after each posture. The score was based on the instructor's judgement of the student progress throughout the training session.

Completion time was recorded from the moment the instructor started the instruction or when the pre-recorded video was played. The training session was considered finished when the student was confident that they could perform the posture. After a pilot testing of the postures, it was decided that 2 minutes was the maximum time required to complete the posture. Should any participant exceeded the 2-minute mark, the research would request a demonstration of the posture for data capture and a maximum completion time of 2 minute would be recorded for that posture.

After completing the 16 postures, the student participant was asked to complete a questionnaire rating their experience with the 4 methods. The questionnaire asked the participants

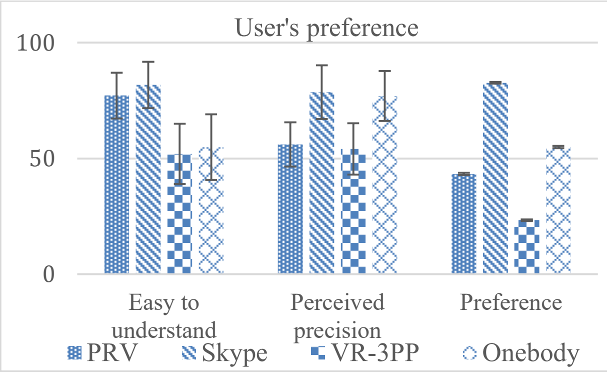

Figure 5. Questionnaire responses from participants

to rate on how easy to understand and the perceived precision of their task, as well as ranking their preference for the methods. Responses to the questions were recorded on a visual analogue scale, where the participants marked a position along a continuous line between two endpoints representing the opposite scales of the response. The scale was converted to a percentage by the researcher for analysis.

Measures were put in place to ensure consistency among the conditions. All errors calculation was performed by the Kinect. For video condition, the error value was compared with a pre-recorded pose by the instructor. The instructor followed the exact same process in delivering the instructions for all conditions. The instructor was recruited to have the ability to demonstrate postures accurately and consistently across all conditions, including the pre-recorded pose for accuracy test with video condition.

Procedure

A researcher greeted the student participant and led them into a living room set up in the lab with a computer, and a 48 inch TV on which a Kinect sensor mounted in front (see Figure 2). In all methods, the student participant performed the postures in front of the Kinect, used to measure the posture accuracy. The researcher provided an instruction for each of the method and remained in the room while the instructor taught the student participant the same sample posture for all methods. The sample posture was not part of the set of 16 experiment postures.

The instructor was set up in a separate room (see Figure 2), standing in front of a 40-inch TV with a Kinect sensor placed in front. The Kinect enables both body tracking and video conferencing, with RGB camera, depth camera, and a microphone.

For the pre-recorded video method, a video was played on loop at the start of the session. For the Skype method, a full screen Skype call was displayed on both TV screens. For VR-3PP and Onebody methods, both the instructor and the student wore an Oculus Rift. An active Skype call is maintained between the living room of the student and the remote room of the instructor throughout the duration of the study to enable verbal communication through all methods (except for the pre-recorded video method where the microphone on the instructor's side would be muted).

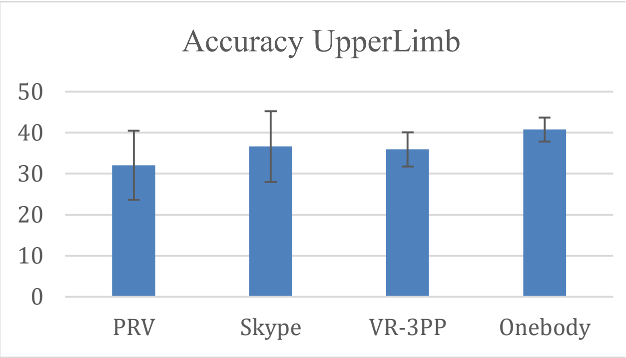

Figure 6. Upper limb accuracy data

The process for each of the training session was as follow:

1. Prepare time: correct video was loaded for PRV. For VR and OB, the student put on the Oculus Rift. Those time were not included in the completion time.
2. Start time: Clock started for completion time. Video started playing on loop for PRV. For all other conditions, the instructor started giving instructions.
3. Progress: the instructions progressed. The order of the instructions was consistent across all methods and postures.
4. Completion: as indicated by the student, both the instructor and the student held the final posture and the research activated Kinect skeletal capturing process. The completion time was recorded when capturing finished.

Results

The qualitative data was checked for outliers. We collected a total of 23 data sets from 23 participants. However, one participant did not complete all conditions due to a software error, therefore, we removed 1 dataset and performed analysis on data from 22 participants.

A scatter plot chart was created for the posture accuracy data of each limb and it was noticed that there was a significantly large number of outliers (accuracy of 0%) for lower limbs data. There was 10% of 0% accuracy in the lower limbs data (3% for upper limb). Verification was performed with recorded video and there were enough similarities between the instructor and the student postures to rule out the possibility of such a large percentage of 0% accuracy. We concluded that tracking errors were significantly large for lower limb data and could not be corrected. Therefore, we decided to discard the lower limb data for accuracy analysis. Pfister et al. [10] performed a comparative study between the Kinect and a motion capture system and also found large error rate using the Kinect for lower body tracking.

We performed one-way within subject ANOVA on the dependent variables of *posture accuracy, completion time,* and *instructor's score* across 4 methods. When there was a significant effect, a post-hoc pairwise comparisons using T-

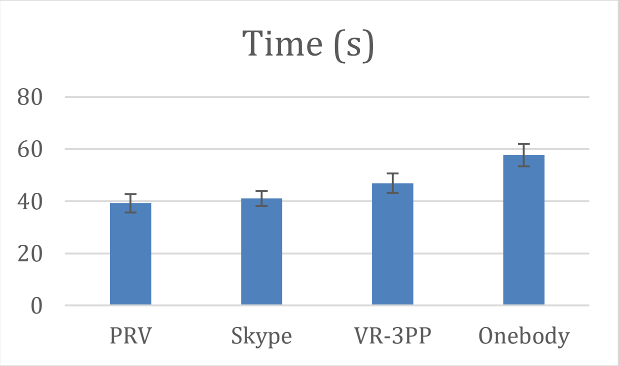

Figure 7. **Time data**

test with Bonferroni corrections were used to find significant differences between the pairs of methods, with the significance level set at $p<0.0125$. The following pairs were compared to highlight the effect of different factors (outlined in 0): PRV and Skype (the factor of synchronous interaction), Skype and VR-3PP (the factor of VR medium), VR-3PP and Onebody (the factor of 1st person view), and Skype and Onebody (the combined factor of VR and 1st person view). Analysis was performed on the mean value across 4 classes of postures.

Accuracy

There was a significant effect ($p<0.05$) on the accuracy value of upper limbs for the four methods [F(3,84)=23.3, p=4E-11). The mean value for Onebody (40.74, SD 5.8) has a significant advantage over Skype (36.6, SD 17.1; $\alpha<0.0125$ and VR-3PP (35.93, SD 8.3; $\alpha<0.0125$). *H1 is supported that Onebody delivers better posture accuracy for upper limbs than existing methods.* There was no significant result found between Skype and VR-3PP. There was a significant advantage of Skype compared to PRV ($\alpha<0.0125$) in terms of upper limb posture accuracy. It must be noted that the accuracy of upper limb has the maximum limit of 50% (25% weighting for each upper limb). Mean and SD values for accuracy are summarised in Figure 6.

Completion Time

There is a significant effect on the measured time ($p < 0.05$) for the four methods [F(3, 84)=28.61, p=7E-13]. Pairwise t-tests indicated that the mean completion time using Onebody (57.7s, SD 73) was significantly higher than Skype (41.1, SD 32.6, $\alpha<0.0125$) and VR-3PP (46.96, SD 57.4, $\alpha<0.0125$). There was a significant effect between Skype and VR-3PP. Figure 7 shows a summary of completion time data.

Instructor's score

There was a significant effect ($p<0.05$) on the instructor's score value for the four methods [F(3,84)=4.64, p=4E-3]. No significant difference was found between the t-test pairs of Skype and VR-3PP, VR-3PP and Onebody, and Skype and Onebody. A significant difference was found between the PRV (mean 7.6, SD 1.22) and Skype (mean 8.5, SD 0.9) methods.

Ease to understand

There was a significant effect on the understanding for the student ($p<0.05$) for the 4 methods [F(3,88)=8.51, p=5E-05]. The mean opinion to understand Onebody (mean 54.87, SD 28.39) is significantly harder ($\alpha<0.0125$) compared to PRV (mean 77.17, SD 19.85). Understanding the instructions using Onebody is significantly harder compared to Skype (mean 81.74, SD 20.02). However, using VR-3PP (mean 52.04, SD 25.99) is not significantly different ($p=0.55>0.0125$) in understanding difficulty from Onebody. Questionnaire responses from the participants are summarized in Figure 5).

Perceived precision

There is a significant effect ($p < 0.05$) on perceived precision [F(3,88)=8.51, p=5E-5], indicating participants found Onebody more precise. For mean opinion in easy to perceive precision shows that using Onebody (mean 76.96, SD 21.63) is significantly easier ($\alpha<0.0125$) to perceive compared to PRV (mean 56.09, SD 19.16). Perceiving precision using Onebody is significantly easier ($\alpha<0.0125$) compared to VR-3PP (mean 54.13, SD 22.19). However, perceiving precision using Onebody is not significantly easier ($\alpha>0.0125$) compared to Skype (mean 78.61, SD 23.16).

Preference

We performed a binomial test on preference data. Overall, 56% of the participant ranked Skype as their most preferred method; 26% for Onebody and 17% for video. No participant ranked VR-3PP as their most preferred method. The preference for Skype is significantly different as the most preferred method ($p=0.0006<0.05$).

DISCUSSION

Higher posture accuracy

Onebody provides better posture accuracy for upper limb instructions. Onebody was significantly better than Skype and VR-3PP, while there was no significant difference between Skype and VR-3PP. This provides an indication that *the main benefit of Onebody for posture accuracy is the first person perspective view*. Compared to the previous study by Chua et al. [11], our study proved a significant advantage of first person view in terms of posture accuracy. Our findings are similar with [2] and [7] in that virtual reality did not present a significant advantage in posture accuracy as compared to video training. It should be noted that their VR implementations are large screen instead of head mounted display. With reference to the factors in Table 1, the study results concluded that *synchronous training (PRV and Skype pair) and first person view (OB and VR-3PP pair) have a positive effect on posture accuracy.*

Onebody allows the user to do things that is not otherwise possible in a physical context. The concept of stepping into the instructor's body and visualizing the instructions from the instructor's own eyes provides a means of bodily communication that is more effective than verbal (Skype) and mirroring view (VR-3PP).

Longer completion time

Onebody took the longest time among all methods. Based on the experience of the researcher observing the study, this did not reflect on the performance of the system. It was observed in the study that using Onebody method, there were more opportunities for the instructor to provide feedback to assist the student. As Onebody uses a matching visualisation that would change the colour of the virtual stick figure when the instructor and the student matched their posture within a certain threshold, both of them spent more time to perfect the posture. In other words, *the visualisation provided by Onebody enables more effective bi-directional feedback between the instructor and the student.*

Participants preference

For each question in the questionnaire, the participant was asked to provide comments. There is some notable mentioning of the issues with leg tracking where the participants expressed difficulty in matching the virtual legs. This observation was in line with our issue with removing lower limb accuracy data due to tracking errors.

The participants were asked to rank their previous experience with VR technology. The mean score for VR experience is 1.77 out of 10 (SD 0.34). Many participants reported to be uncomfortable with VR technology. Some comments included “VR (was) too vague”, and “(I) did not like the stick figure visual and prefer to see the full body (video)”. Issues with visual and proprioceptive conflict was also reported when one participant saw a small discrepancy in the virtual arm’s position to his proprioceptive feeling of where their actual arm was in space. The unfamiliarity with VR technology in a majority of participants may give an indication as to why Onebody did not received high preference ranking, as compared to more traditional method such as Skype. It should also be noted that Skype is a ubiquitous technology. It was expected in the study that participants were very comfortable and familiar with the technology. Therefore, while the participants received training for all the methods equally, Skype would have had an inherent advantage over Onebody and VR methods, which reflected in the participant’s preferences and their completion time. On the other hand, there were several positive comments on how easy it was to follow the virtual avatar in Onebody with one participant stating that Onebody combines the strength of all other methods.

LIMITATIONS AND FUTURE WORK

The study findings highlighted some limitations of our Onebody prototype system.

Level of details

Onebody utilises a simple stick figure as the virtual avatars. The joints are represented by a sphere and the body segments are cylinders, see Figure 1A. Other VR systems [2, 11] implemented virtual realistic humanoid avatars. It is interesting that the usage of a simplified stick figure avatar did not affect negatively on the accuracy result. In fact, we believe that the stick figure avatar provided a simpler matching visualisation as concurrent feedback, as shown in Figure 1. This could have a potential advantage on the cognitive load during training. However, our study did not measure any effects in this aspects. Future study can investigate the effects of cognitive load as well as implementation using more realistic humanoid avatars.

Avatar superimposition

Onebody attaches the avatars of the instructor and students at the hip to correctly superimpose the avatars. While this approach is effective for martial arts teaching, as demonstrated by our experiment, it may not be applicable for other motor skill training areas, such as dance, yoga, or medical therapy skill training. A more systematic approach to overlay rendering is needed to adapt Onebody for different motor skill training domains.

Posture, not movement

The initial implementation of our system targets posture training for martial arts. We implemented comparison metrics between the 2 static avatars to provide concurrent feedback on posture accuracy. Our approach is not directly applicable for dynamic movement guidance. Future research in this direction will expand the applicability of Onebody.

Application areas

While initially aimed at martial arts training, Onebody has the potential to benefit other areas of martial arts training. Similar domains such as dance, yoga, or sports training could benefit from our system. Manual therapy education is an area where augmented feedback is vitally important yet challenging. In a typical postures class with a teach to student ratio of 1:20, the tutors cannot provide adequate feedback to all the students, in performing and understanding the basic principles and skills needed to effectively treat patients with pain and disabilities. We are interested in exploring the potential of augmented feedback using Onebody in those application areas.

CONCLUSION

Onebody is a virtual reality remote posture guidance system that uses first person perspective. The application for our system is within sports or physical activity training, such as martial arts, yoga or dance. The system implements skeletal tracking of an instructor and the student, rendered as overlaid avatars. Using a virtual reality headset, the student can visualise the movements of the instructor as if they were the instructor. Onebody provides a first person view of the instructions, allowing the student to *step into the instructor’s body*. We conducted a study to compare multiple aspects of Onebody against existing techniques. The aspects include synchronous interaction, VR training, and first person perspective. Compared to previous work, our study combined those aspects into a single experiment and proved significant advantage of the first person view versus third person view in virtual reality, in terms of posture accuracy.

REFERENCES

1. Anderson, F, Grossman, T, Matejka, J, and Fitzmaurice, G, YouMove: enhancing movement training with an augmented reality mirror, in *Proceedings of the 26th annual ACM symposium on User interface software and technology*. 2013, ACM: St. Andrews, Scotland, United Kingdom. p. 311-320.
2. Bailenson, J, Patel, K, Nielsen, A, Bajscy, R, Jung, S-H, and Kurillo, G, The Effect of Interactivity on Learning Physical Actions in Virtual Reality. M*edia Psychology,* 2008. 11(3): p. 354-376.
3. Chan, JCP, Leung, H, Tang, JKT, and Komura, T, A Virtual Reality Dance Training System Using Motion Capture Technology. *IEEE Transactions on Learning Technologies,* 2011. 4(2): p. 187-195.
4. Chinthammit, W, Merritt, T, Pedersen, S, Williams, A, Visentin, D, Rowe, R, and Furness, T, Ghostman: augmented reality application for telerehabilitation and remote instruction of a novel motor skill. *BioMed research international*, 2014. 2014.
5. Covaci, A, Olivier, A-H, and Multon, F, Third person view and guidance for more natural motor behaviour in immersive basketball playing, in *Proceedings of the 20th ACM Symposium on Virtual Reality Software and Technology*. 2014, ACM: Edinburgh, Scotland. p. 55-64.
6. Holden, MK, Virtual Environments for Motor Rehabilitation: Review. *CyberPsychology & Behavior,* 2005. 8(3): p. 187-211.
7. Jung, S-H and Bajcsy, R, A framework for constructing real-time immersive environments for training physical activities. *Journal of Multimedia,* 2006. 1(7): p. 9-17.
8. Kallmann, M, Camporesi, C, and Han, J, VR-Assisted Physical Rehabilitation: Adapting to the Needs of Therapists and Patients, in *Virtual Realities.* 2015, Springer. p. 147-168.
9. Kurillo, G, Bajcsy, R, Nahrsted, K, and Kreylos, O, Immersive 3D Environment for Remote *Collaboration and Training of Physical Activities, in Virtual Reality Conference,* 2008. VR '08. IEEE. 2008. p. 269-270.
10. Pfister, A, West, AM, Bronner, S, and Noah, JA, Comparative abilities of Microsoft Kinect and Vicon 3D motion capture for gait analysis. *Journal of Medical Engineering & Technology*, 2014. 38(5): p. 274-280.
11. Philo Tan, C, Crivella, R, Daly, B, Ning, H, Schaaf, R, Ventura, D, Camill, T, Hodgins, J, and Pausch, R, Training for physical tasks in virtual environments: Tai Chi, in *Virtual Reality, 2003. Proceedings. IEEE*. 2003. p. 87-94.
12. Schmidt, RA and Wulf, G, Continuous concurrent feedback degrades skill learning: Implications for training and simulation. *Human factors,* 1997. 39(4): p. 509.
13. Sigrist, R, Rauter, G, Riener, R, and Wolf, P, Augmented visual, auditory, haptic, and multimodal feedback in motor learning: a review. *Psychonomic bulletin & review,* 2013. 20(1): p. 21-53.
14. Tang, R, Yang, X-D, Bateman, S, Jorge, J, and Tang, A, Physio@ Home: Exploring visual guidance and feedback techniques for physiotherapy exercises, in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems.* 2015, ACM. p. 4123-4132.
15. Tao, G, Archambault, PS, and Levin, M, Evaluation of Kinect skeletal tracking in a virtual reality rehabilitation system for upper limb hemiparesis, in *Virtual Rehabilitation (ICVR), 2013 International Conference on. 2013, IEEE.* p. 164-165.
16. Todorov, E, Shadmehr, R, and Bizzi, E, Augmented Feedback Presented in a Virtual Environment Accelerates Learning of a Difficult Motor Task. *Journal of Motor Behavior,* 1997. 29(2): p. 147-158.
17. Wulf, G, Shea, C, and Lewthwaite, R, Motor skill learning and performance: a review of influential factors. *Medical Education,* 2010. 44(1): p. 75-84.
18. Wulf, G and Shea, CH, Principles derived from the study of simple skills do not generalize to complex skill learning. *Psychonomic Bulletin & Review,* 2002. 9(2): p. 185-211.
19. Wulf, G and Shea, CH, Understanding the role of augmented feedback. *Skill acquisition in sport: Research, theory and practice,* 2004: p. 121-144.
20. Wulf, G and Weigelt, C, Instructions about physical principles in learning a complex motor skill: To tell or not to tell…. *Research quarterly for exercise and sport,* 1997. 68(4): p. 362-367.
21. Yang, U and Kim, GJ, Implementation and Evaluation of “Just Follow Me”: An Immersive, VR-Based, Motion-Training System. Presence: *Teleoperators and Virtual Environments,* 2002. 11(3): p. 304-323.