

YouMove: Enhancing Movement Training with an Augmented Reality Mirror

Fraser Anderson^{1,2}, Tovi Grossman¹, Justin Matejka¹, George Fitzmaurice¹

¹Autodesk Research, Toronto, ON, Canada
{firstname.lastname}@autodesk.com

²University of Alberta, Edmonton, AB, Canada
frasera@ualberta.ca

ABSTRACT

YouMove is a novel system that allows users to record and learn physical movement sequences. The recording system is designed to be simple, allowing anyone to create and share training content. The training system uses recorded data to train the user using a large-scale augmented reality mirror. The system trains the user through a series of stages that gradually reduce the user's reliance on guidance and feedback. This paper discusses the design and implementation of YouMove and its interactive mirror. We also present a user study in which YouMove was shown to improve learning and short-term retention by a factor of 2 compared to a traditional video demonstration.

Author Keywords

Movement; full body; 3D; learning; training; augmented reality; motor learning; guidance.

ACM Classification Keywords

H.5.2 [Information interfaces and presentation]: User Interfaces. - Graphical user interfaces.

INTRODUCTION

Mastering new postures and motions is a crucial component of many physical activities such as dancing, martial arts, and sports. Learning these motor skills can be challenging, requiring hours of training and repetitive practice [38]. Often, learning involves enrolling in classes where an instructor leads a group through a set of exercises. Ballet dancers often use mirrors to receive visual feedback in addition to coaching [7]. While there may be no replacement for expert coaching, less formal self-paced learning may be more desirable for in-home practice, to supplement professional coaching, or for hobbyists.

Technology has enabled movement training to occur at home. Prior to the Internet, trainees could practice along with instructional VHS tapes, such as exercise videos. Now, online video sites, such as YouTube, offer thousands of movement training videos, for skills ranging from surgery to jiu-jitsu. While videos for a variety of skills are available, the use of video as a training medium is limited. Videos offer no feedback, do not capture 3D movement information, and offer no personalized motivation.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.
UIST'13, October 08 – 11, 2013, St Andrews, United Kingdom.
Copyright © 2013 ACM 978-1-4503-2268-3/13/10...\$15.00.
<http://dx.doi.org/10.1145/2501988.250204>

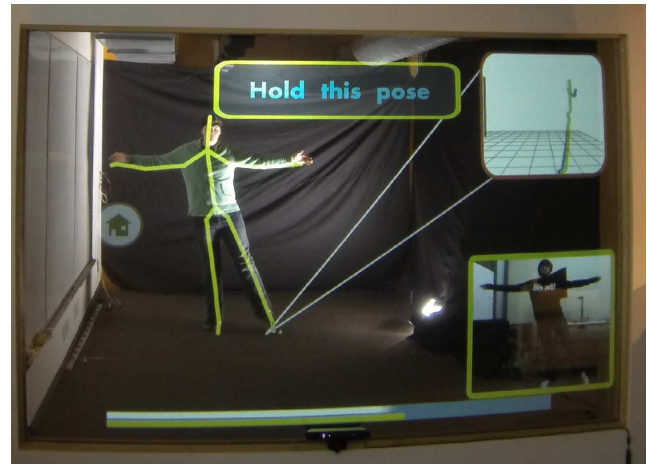


Figure 1: YouMove allows users to record and learn physical movement sequences. An augmented reality mirror provides graphic overlays for guidance and feedback. Note that for this photo the virtual viewpoint was vertically repositioned to account for the offset of a head-mounted camera, and floor lighting was used to reduce glare.

There has also recently been an emergence of fitness and dance video games that use game consoles with tracking technology such as the Microsoft Kinect [6]. While these games may improve some movement skill, their target is entertainment and exercise. In these games, movements are constrained to pre-programmed sequences, and the feedback is limited, often consisting of a simple 2D outline.

In this paper, we present YouMove (Figure 1), a system for learning full body movements. YouMove is comprised of a Kinect-based recording system, and a corresponding training system. The recording system is designed to be easy to use, so anyone can capture movement sequences and annotate them for learning, without the need for complicated motion capture hardware or software. The training system uses the recorded video and 3D movement data to guide the trainee through a series of interactive stages. The training system augments a traditional 'ballet mirror' experience by using a half-silvered mirror with graphic overlays for guidance and feedback. The use of a mirror allows for zero latency, high fidelity feedback. In a user study we find YouMove improves movement learning and short-term retention by a factor of 2 compared to traditional video demonstrations. As implemented, YouMove is immediately beneficial for several domains of movement, such as yoga, dance, and physical therapy, with more domains possible with further advances in sensing.

This paper provides several contributions. First, we provide the first generalized full-body movement training system, YouMove. Next, we provide design guidelines, implementation details, and interaction techniques for a whole-body, interactive, augmented reality mirror. Third, we provide an evaluation demonstrating YouMove's effectiveness over a traditional video approach. Last, we analyze learning and preferences with YouMove and provide directions for future work.

RELATED WORK

This work is influenced by prior work in motor learning, technology-mediated skill learning, and display technology.

Motor Learning

Within the domain of motor learning, there is substantial research on the effects of various factors on motor skill acquisition. The amount of deliberate practice has been shown to have the greatest effect on learning [38]. However, several factors influence the effectiveness of the practice. In particular, the availability and modality of feedback can greatly impact skill acquisition [30].

Nearly all training systems and research involve various manipulations of feedback, see Sigrist et al. for a thorough review [30]. Audio has been shown to be beneficial in learning timing [28]. Summary feedback (feedback at the end of a series of trials) can improve learning compared to feedback shown after each trial [26]. Allowing learners to control when they see feedback is also beneficial [13].

Technology-Based Motor Skill Training

Virtual reality has been leveraged for training movement in a number of specific domains. Various surgical simulators have shown promise in improving surgical skill [14]. A virtual ping-pong trainer showed positive effects of training [32]. The use of a virtual 'ghost' for training 3D hand movements showed that a VR environment was no worse than real-world training [43]. Several other VR-based trainers have had similar results showing a custom system performing at least as well as a traditional method [16]. In each of these cases, the system is designed to support very specific movements, leading to a lack of generalizability.

Physical therapy has driven the use of technology for movement learning [12]. These typically focus on re-training simple movements and improving range of motion. A number of systems have been developed around gaming and VR platforms in an effort to make the otherwise monotonous or painful movements motivating [40]. There is evidence suggesting that the major benefits of technology-based therapy tools are in their ability to motivate and encourage adherence to training schedules [1].

Within the HCI literature, several systems for learning gestures for computer input have been developed and studied. Octopocus provides a dynamic, real-time guide for learning 2D stroke gestures [3]. ShadowGuides provides a similar guide for learning multi-touch gestures and hand postures [9]. Users of both systems showed improvement in

learning the mapping of gestures. Anderson and Bischof studied various gesture guides and found high guidance can hinder learning of the motor component [2].

Additional work within the HCI community has focused on movement guidance. Henderson and Feiner developed an HMD-based augmented reality system providing real-time guidance for assembly tasks [11]. LightGuide used on-body projections to guide a user's hand through 3D space [31]. With this approach, feedback is only available for a single hand, and only when the hand is visible. ModelMA allows a performer to record a repeated movement using the Kinect, which an observer can then practice with, however, learning is not addressed and is left as future work [35]. These systems were limited to guiding a small set of movements. In contrast, YouMove supports full-body movement learning using interactive, gamified training.

Novel Display Technologies

Half-silvered mirrors (HSMs) have been used in a number of interactive scenarios. The i-mirror used a projector to allow users to apply virtual makeup, and provided artificial 'night vision' on a small display [33]. The aware-mirror and MR-mirror use an LCD display behind a HSM, displaying news and information, or playing games [10, 25]. In the Haunted Mansion at Walt Disney World, a large display is placed equidistant from the viewer, opposite a HSM, allowing the viewer's reflection and the virtual content to be displayed on the same focal plane [8]. The Hologlector used the same technique coupled with a Kinect to enable 'holographic' interaction with virtual content [19]. This configuration allows reflection-aligned content from multiple viewpoints, but requires a substantially larger configuration space and requires users to maintain a fixed distance from the screen.

Our work is also guided by research on free-hand interaction with large displays. VideoPlace allowed users to visualize their body movement on a large projection screen [15]. Shadow reaching extended this concept by allowing users to use their shadow to select targets that were previously out of reach [29]. Several works have also explored methods for selection on large displays [22, 36].

INSIGHTS ON TRADITIONAL INSTRUCTION

To gain insights that could inform the design of YouMove, we observed professional yoga and ballet classes.

Our observations were consistent with the strategies discussed in the recent work of Velloso et al. [35]. We observed a feedback loop between trainer and trainee consisting of demonstration, performance, and feedback. For example, both the ballet and yoga instructors would demonstrate the movement sequence, then have the trainees repeat the sequence as they made comments and physical corrections to the trainees' movements.

We also observed the ballet instructor using an adaptive guidance approach, reducing the amount of guidance as the class progressed. The trainees started at the barre, then

moved to the mirror, and finished the class with more complex sequences without the use of the mirror. In both classes instructors provided timing cues, motivation, and alternative metaphors for thinking about the movement.

DESIGN GUIDELINES

Based on our observations of instruction strategies and a review of previous literature, we have developed a set of design goals to guide the development of the YouMove system. We believe these guidelines are applicable to movement training systems in general.

Leverage Domain Knowledge

Experienced trainers have more knowledge than what can be expressed in a recorded movement. Our observations indicated that feedback from instructors is based on domain knowledge. The experts know how to segment movement, and which parameters of the movement are important. The authoring system should support the capture of this domain knowledge with minimal effort.

Motivate the User

Engagement and motivation is an important part of learning [4]. The training system should provide feedback on the user's progress, encourage the user to continue with their practice, and make it an enjoyable experience.

Simple Presentation, Low Cognitive Load

While practicing, the users should maintain a low cognitive load [34]. The user's attention should be on the movement, not on interpreting complex UI elements. A direct representation of the movement and simple scoring measures should focus the user on learning the movement.

Adaptive guidance

Excessive guidance and video demonstrations can hurt learning, as users come to rely on the guides [21, 27]. Guidance should be reduced as the user learns

Summary Feedback

Feedback on individual movements can cause trainees to focus on small errors, rather than the larger systemic errors. Aggregating feedback from several performances allows trainees to see systemic errors in their movement [39].

User driven learning

Trainees have varying skill levels and preferences that will dictate their training needs. The system should allow users to progress at their own pace to maximize learning [41].

AUGMENTED REALITY MIRROR IMPLEMENTATION

While there have been previous implementations of virtually enhanced physical mirrors [8, 19], we are unaware of a technical setup similar to our own in the research literature. This section contributes a technical description of our Augmented Reality Mirror setup, which could have implications that go beyond the YouMove system.

Our use of the mirror for YouMove is inspired by the floor-to-ceiling mirrors often found in ballet studios. These mirrors allow dancers to see their movements, providing them with immediate visual feedback [7]. By augmenting a

traditional mirror with interactive content, we provide additional information that helps trainees learn a movement.

Configuration

The configuration of the Augmented Reality Mirror is illustrated in Figure 2. The display consists of a 3.2m x 1.8m pane of glass with a half-silvered mirror film applied to one side, and a diffuse film applied to the other. The mirror film¹ transmits 16% of the visible light, and reflects 58%, resulting in a highly reflective surface while still allowing projected light to pass through. Several other mirror films were tested, but the others absorbed more of the projector light, or did not reflect the user's image as clearly. The diffuse film serves to diffuse the light from a rear-mounted projector (Mitsubishi FD630U, 1920x1080 pixels).

A Microsoft Kinect is mounted below the mirror to track the user. The location of the mirror is specified in the coordinate space of the Kinect. The position of the user's head and the corners of the mirror define an asymmetric projection matrix used to render on-screen content.

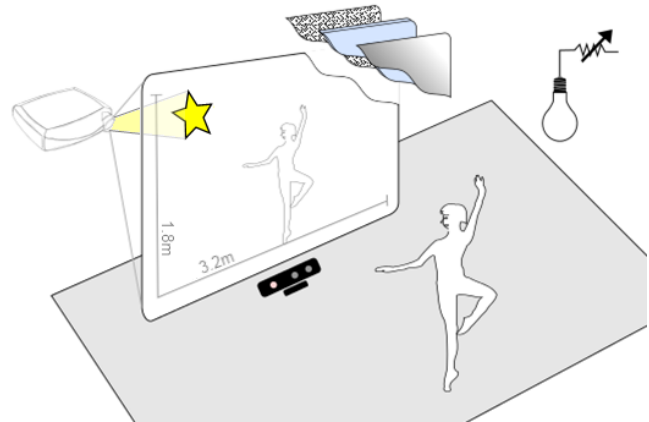


Figure 2: Overview of system design showing projector, layered screen, dynamic lighting, Kinect and user location.

Lighting

While the system is usable in varying ambient lighting conditions, the experience is improved by being in an environment with controlled lighting. With bright ambient light and a dark projection image, the user focuses solely on their reflection in the mirror. With low ambient light and a bright projection, the user's reflection disappears. With a moderate amount of ambient light, the user's reflection and the projected image are both visible (Figure 1). The ability to control lighting allows the system to shift the user's attention between the relevant images (virtual or reflected).

To manipulate the ambient light, a servo motor was mounted to a light's dimmer switch and connected to the PC through an Arduino over USB. The servo motor was mounted using modular building blocks (LEGO), allowing

¹ Supreme Silver 20 – www.apexfilms.ca

the motor to be easily removed and accurately replaced later. The physical actuation of existing dimmer switches would not require the end user to modify any existing wiring or manually adjust the lighting.

Interaction

Mirror-based augmented reality offers unique opportunities for interaction. The user's reflection can be used to directly activate on-screen components, allowing for direct manipulation of a 2D interface from 3D free-space. This *reflection selection* provides zero latency feedback on hand position, allowing quick positioning [20]. Buttons are activated by dwelling the hand over the button. During the dwell period, the button expands providing feedback and increasing the activation area in case the user's hand drifts.

We implemented two types of buttons. Global menu buttons are located on the left side of the mirror, so they will not be triggered accidentally during training. The vertical location of the global buttons adapts to the user's height, so that menu items are not out of reach. Quick-access contextual buttons are presented near the user's head, and only appear when required. These buttons are convenient, and their body-centric positioning allows them to be activated by a 'gesture posture', as their location relative to the body is constant, similar to Virtual Shelves [18].

YOUMOVE IMPLEMENTATION

YouMove is composed of a simple program to record data, and a separate training system for playback of that data. The authoring and training system are designed to run independently. The recording software is written in C#, using the WPF framework and the Kinect SDK (v1.6). The training software is written in C++, using openFrameworks, OpenGL, and the Microsoft Kinect SDK v1.6. All software was run on a Window 7 PC, with 12 GB of RAM and a dual core processor. The only calibration required for the training system is to specify the location of the mirror when it is first set up.

YOUMOVE AUTHORIZING SYSTEM

We developed custom software to capture a trainer's movement and domain knowledge. By simplifying capture, YouMove allows experts to contribute learning material without the need for complicated motion capture hardware or software. The recording system allows authors to record themselves performing the movement. The system captures video, audio and 3D skeleton movement data of the author.

Recording

After launching the software, the author is presented with a screen that has a single 'Record' button as well as the current video stream from the Kinect with a skeleton overlay. To capture movement, the author presses record, performs the movement, and then presses the stop button. Pressing stop takes the author to the editing interface.

Editing

The editing interface (Figure 3) allows authors to trim the recording to eliminate unwanted data – such as walking to

or from the capture volume. The author can also specify global parameters for the movement, i.e., timing, smoothness, precision, or stability. These parameters are used by the training system when providing feedback.

Authors then specify keyframes for the recorded movement. Keyframes are postures within the movement that are particularly important for a trainee to match during the movement. For example, a keyframe for a baseball throw may be when the hand reaches peak extension.

Keyframes are specified by navigating to the desired frame and then directly clicking on joints to specify the *important joints* for that keyframe. When a joint is selected, the current frame becomes a keyframe. These keyframes and important joints will be used in the training system to provide tailored guidance and feedback.

Authors can also associate an additional audiovisual recording with individual keyframes, allowing the trainer to provide additional information regarding the movement. This annotation is done by clicking the 'record' and 'stop' buttons below each keyframe marker. Annotations can be used to provide information that may not be immediately obvious to a trainee from seeing the author's movement, or to discuss common pitfalls to avoid. For example, to throw a baseball, an author may annotate the keyframe with a short clip explaining that the elbow should be at approximately 90 degrees.

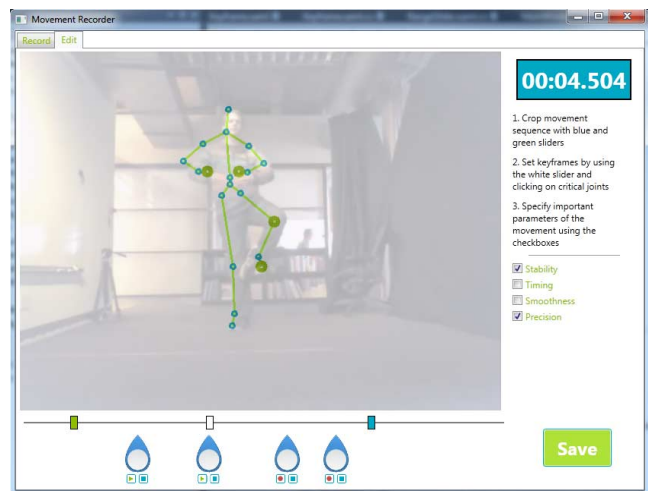


Figure 3: The editing interface allows authors to specify keyframes and global movement parameters. Each keyframe specifies the important joints for that moment, and can be associated with a recording that provides detail.

Saving and Sharing Captures

The user saves the capture by clicking the Save button. The data is saved as media files (mp4 video and wav audio) and plain-text files containing timestamped skeleton locations and keyframe metadata used to synchronize the data.

YOUMOVE TRAINING SYSTEM

The training system is used to teach the movements previously recorded with the authoring system.

Movement Gallery

The initial screen of the system presents a gallery of movements that the user may wish to learn (Figure 4). All buttons and icons throughout the entire system are selected using the reflection selection technique.

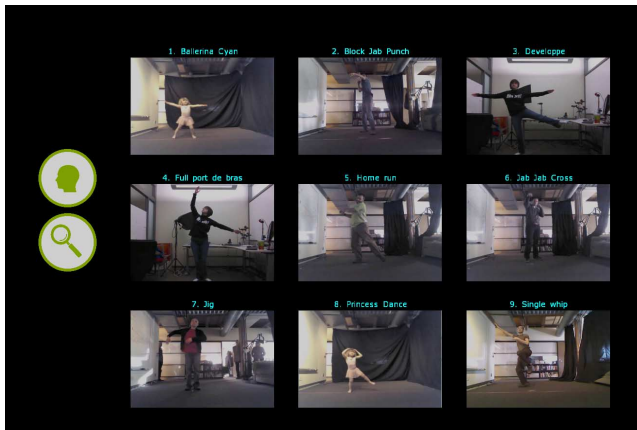


Figure 4: The movement gallery allows users to change profiles, query by example, and select a movement.

Query by Example

Selecting the magnifying glass allows users to search for a movement by example. The search screen instructs users to hold a representative posture of the desired movement. Once the user stays still, the system searches the movements in the library for the best match, presenting the most similar movements in a grid for the user to select.

The posture similarity heuristic used for the search is the same as what is used for the scoring measure used in training. That is, it takes a ‘snapshot’ of the posture the user performing, and compares it to each of the keyframes in each movement of the library. The ability to query by movement may also be useful to search for movements where the movement is emphasized, rather than the pose.

Skeleton Alignment

A fundamental feature of the YouMove training system is guidance and feedback based on a comparison of the author’s and trainee’s skeleton. The Kinect provides skeleton tracking, reported as 20 joints, with 3D positions updated at 30Hz. The tracked joints include large body parts, such as the hands, arms, torso, legs and the head, but fine movements (e.g., the fingers) are not tracked.

To properly calibrate the author’s training skeleton to each user, it must be scaled and translated to match the user’s size and position. This is necessary for the skeleton-driven feedback and for accurate scoring. The spatial alignment is done by aligning the hips of the author skeleton with the hips of the trainee. Orientation is not aligned, as it is assumed that the trainee and trainer will be performing the movements in the same orientation relative to the Kinect.

Scaling the skeleton is achieved by dynamically resizing each bone in the trainee skeleton to match the size of the corresponding bone in the trainer. This scaling is done

hierarchically from the hips to propagate changes throughout the skeleton. This method is necessary as a simple uniform scaling would not accommodate users with proportionally different limb lengths. We have found that this skeleton algorithms works robustly, and allows the system to work well even when the trainee has a significantly different age, height or weight from the author.

Scoring and Stage Progression

The YouMove system incorporates several elements of gamification designed to motivate the user [17]. Training is composed of a series of stages, and each stage scores the user’s performance based on the similarity between their movement and the target movement. If their performance is high enough, they get a gold star and are allowed to progress to the next stage. To avoid frustration, a stage is also unlocked if they repeat the stage twice.

Each keyframe is scored based on the joint with the maximum error, measured by Euclidean distance. Only *important joints* specified in the authoring tool are used to compute the score. To allow for small errors in timing, a window of 0.5s on each side of the target frame is searched to find the best matching posture. If the author has specified that timing is important in the authoring tool, this window is decreased to 0.25s. The maximal Euclidean distance is mapped to a score using a linear mapping, with 0 error being a perfect 10, and 0.15m of error resulting in the score of 7.5 needed to get a gold star and unlock subsequent stages. If precision is specified as a global parameter, this mapping is modified so 0.10m results in a score of 7.5. These values were determined by experimentation.

Training Stages

Once a movement is selected from the gallery, the system progresses through five stages: *Demonstration*, *Posture Guide*, *Movement Guide*, *Mirror*, and *On Your Own*. The user can navigate through the stages using a selection screen (Figure 5, left), although the locking mechanism forces the user to initially perform the stages in order.

The stages progressively introduce the movement to the trainee, and gradually reduce their reliance on guidance and feedback. Each stage presents the user with unique challenges and a different context to perform the movement in, reducing the negative impact of specificity of learning [24]. The Posture Guide repeats twice, and the Movement Guide, Mirror and On Your Own repeat five times, allowing users to practice without being interrupted.

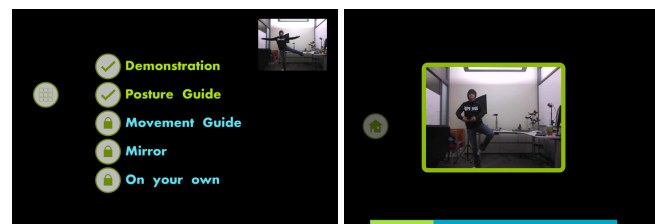


Figure 5: Left) Stage selection interface allowing users to begin one of the unlocked stages. Right) Demonstration stage.

Demonstration

The demonstration stage plays the recorded video for the trainee (Figure 5, right). This stage is designed to be simple and familiar, so the focus is on the movement to be learned.

Audio is played alongside the video to help with timing. At the start of the movement, a pre-recorded voice speaks the word ‘And’, and each keyframe is counted out in sequence, i.e., ‘One’, ‘Two’, etc. The use of numbers (rather than a metronome click) helps trainees remember each movement in the sequence. This counting is present in all other guides except for the ‘On Your Own’ guide. A progress bar on the bottom of the screen is synched to the video playback.

Posture Guide

The posture guide (Figure 6) helps trainees refine their body position by pausing the movement at each keyframe and providing real-time feedback on their errors. This stage is inspired by the tutorial system found in Pause and Play [23], as it halts the tutorial and allows trainees to work at their own pace. The stage pauses until the user holds a stable position for one second, or until five seconds have elapsed. An additional progress bar on the bottom of the screen represents the amount of time they have currently been stable. If stability is specified as a global parameter in the authoring tool, the threshold for detecting stable movement is lowered from 5 to 3 cm per frame.

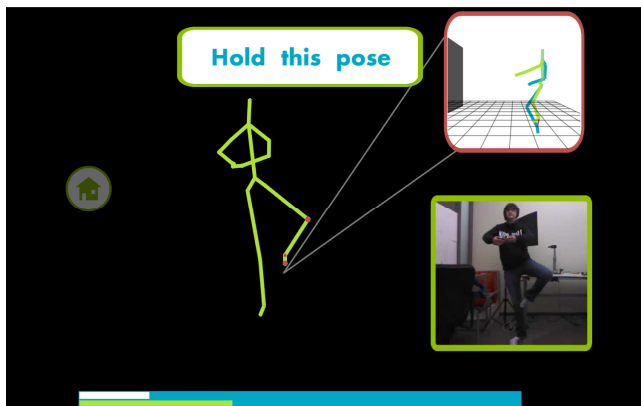


Figure 6: The posture guide requires trainees to maintain a stable posture, matching the position of the trainer. Errors in joint position are indicated by red circles. The callout (top right of figure) can be used to resolve depth ambiguities.

In this stage, ambient lighting is increased and the trainer’s skeleton is virtually aligned to the trainee on each rendered frame. Trainees are instructed to match their reflection with the trainer’s green skeleton, as in Figure 1. Errors in positioning are shown as red circles overlaid on the user’s joints, with the circle radius proportional to the joint error. If the largest error is found to be in the z-axis (depth), then a call-out window appears that shows the trainee a side-view of their body position superimposed on the trainer movement. The author’s original video is also displayed on the screen for additional visual reference. The video is automatically cropped based on the location of the trainer’s skeleton within the video to preserve on-screen space.

Movement guide

The movement guide (Figure 7) provides a skeleton similar to that of the posture guide, but it moves in real-time and does not pause on the keyframes. In addition, the trainer’s skeleton is aligned relative to the starting location of the trainee on each repetition – forcing the trainee to move with the skeleton to maintain alignment.

The movement guide also provides ‘cue ribbons’ – 3D trajectories that help the user with timing by visualizing the upcoming movements. The ribbons display the trajectory of the hands and feet 300ms ahead of the current frame, with the ribbon becoming more opaque the farther it is in the future. To reduce visual complexity, the ribbons are only displayed if the joint is going to surpass a velocity threshold of 75 cm/s within the subsequent 300 ms. If smoothness is specified as an important parameter, the ribbons are extended to movement 500ms in the future, to allow trainees to better prepare their upcoming trajectories.

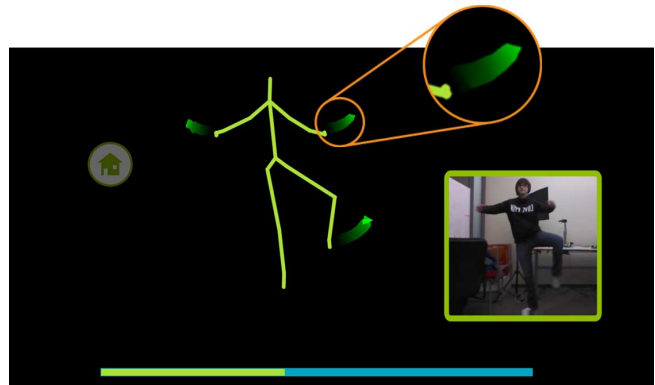


Figure 7: The movement guide encourages the trainee to match the trainer’s movements at full speed, using green ribbons (inset) to cue upcoming movements.

Mirror guide

The mirror guide encourages the trainee to focus on their reflection, and does not provide any visual cues to guide the movements. A black screen is projected onto the mirror, and the ambient lighting is increased to enable the trainee to see a clear reflection of themselves in the mirror. Audio cues are present to help with timing. This type of guide is similar to a ballet class, where the student practices in front of the mirror with the instructor counting the beats.

On your own

The ‘On Your Own’ guide mimics a real-world performance scenario where the trainee relies only on what they have learned. The ambient lighting is off, and a white image is projected onto the mirror, preventing the trainee from seeing any reflection. The only audio cue provided is the word ‘and’, used to indicate the start of the movement.

Post-Stage Feedback

Trainees can view their performance using the Post-Stage Feedback screen (Figure 8). This screen is displayed after each of the training stages, except the Demonstration. This screen summarizes the each of the previous repetitions and

presents the aggregate data for each keyframe. Trainees can navigate between keyframes to see an average score for that keyframe, as well as their body pose and a summary video for that frame. During Post-Stage Feedback, ambient lighting is kept at a moderate level, and the projected background is grey. This enables the trainee to see the on-screen content, while allowing them to see their reflection and use reflection-based buttons.

Error in posture is represented by two skeletons: the average skeleton of the trainee for all repetitions (blue), and the trainer skeleton (green). The same circles used in the posture guide are used for feedback to indicate relative joint error. While viewing the feedback, trainees can rotate the 3D view of the skeletons by walking left or right, allowing them to see errors in all dimensions.

A summary video is also provided for each keyframe, allowing the trainee to quickly assess their movements and compare them to the trainer. The trainer video is a static image taken from the recorded video, while the trainee video is an animated sequence of images, each taken from one repetition of the movement, displayed for 0.5s each. By animating the images, trainees can easily see the variations in their movement from frame to frame.

A group of 5 contextual buttons are displayed around the user, allowing them to re-perform the previous stage, advance to the next stage, navigate between keyframes, or view a video annotation if one has been associated with the current keyframe.

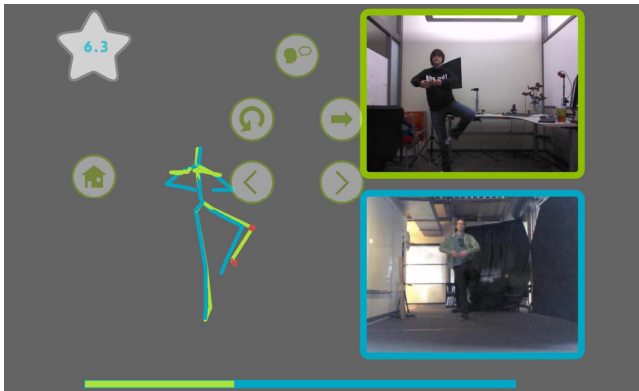


Figure 8: The Post-Stage Feedback screen, showing trainees their overall score, average position (skeleton), and video for each keyframe.

EVALUATION

A controlled study was conducted to compare YouMove to traditional video-based instruction methods. Effectiveness of movement instruction was compared using the results of retention scores after training with each system. Subjects also provided qualitative feedback on the YouMove system, and the use of a mirror as an interactive display.

Participants

Eight participants (2 female) between 21 and 51 years old, (\bar{x} = 30.1 years) were recruited from within the

organization, but external to the group. No participants had prior knowledge of, or experience with the system or study, or a dance background.

Study Movements

An author of this paper recorded four movements for the study – two ballet movements and two abstract movements. The ballet movements (variations on the *Tendu* and *Develope*) were easier to conceptualize and only required a moderate amount of movement. The abstract movements were more difficult to perform, as they were a series of postures with no clear structure and required substantial movement. All movements used 4 keyframes.

Conditions

The two conditions for the study were *YouMove* and *Video*. The *YouMove* condition consisted of the *YouMove* system as previously described. None of the movements contained multimedia annotations or had global movement parameters specified. The *Video* condition consisted of a video being projected on the mirror screen at the same size as the *Demonstration* guide (102 x 76 cm), with the area on the screen around the video projected white to prevent the participants from seeing their reflection.

Design

The study was conducted as a two factor repeated-measures design, with each participant learning two movements using *YouMove* (one abstract followed by one ballet), and two by using just the demonstration video. The condition order and movement pair to condition mapping was fully counterbalanced across the 8 participants.

Procedure

Each participant was introduced to the *YouMove* system using a simple tutorial movement. The experiment began once participants were comfortable with the system.

Each movement began with a pre-test phase, in which the participant watched a video of the movement twice, and then performed the movement five times during the ‘On Your Own’ guide. Next, the participant trained using either the full *YouMove* system, or just the *Video*. During training, the participant practiced 45 times, either by practicing along with the video, or by practicing with the various guides of the *YouMove* system. In the *YouMove* condition, users were free to navigate to any of the unlocked training stages.

Participants completed a short-term retention test five minutes after completing the training. The test consisted of watching the demonstration video once to review the movement, then performing the movement five times during the ‘On Your Own’ guide. The Kinect recorded the participant’s movements in all stages of the study.

A short questionnaire was given to participants after completing the tasks to elicit qualitative feedback. The study took approximately two hours to complete.

Results

A repeated measures ANOVA with two independent variables: *condition* (YouMove, Video), and *movement type* (Ballet, Abstract) was performed. Performance (Figure 9) was measured by computing the RMSE between the space-and-time aligned user skeletons and the target skeleton, using the neck, hands, elbows, knees and feet. These joints were chosen because they are reliably tracked and capture the majority of the information in the movements. Similar analyses were conducted using only the joints specified by the content author, as well as using the joint that has maximal error (the measure used in computing the in-game score). All of these analyses produced equivalent statistical findings, and so we only present the RMSE analysis.

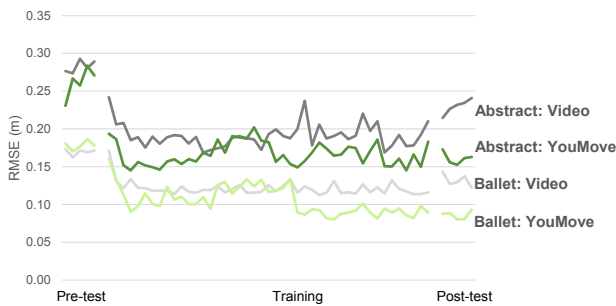


Figure 9: Performance on each trial, for each of the conditions, averaged over all 8 subjects.

The change from pre-test to post-test was significant ($F_{(1,7)} = 9.98, p = 0.02$), with YouMove scores improving by an average of 0.10m (44%), and the Video condition improving scores by 0.05m (20%), representing a medium-large effect size ($\eta^2 = 0.13$). The effect of movement type on the change was not significant ($F_{(1,7)} = 0.24, p > 0.6$), nor was the interaction between movement type and condition ($F_{(1,7)} = 0.23, p > 0.7$), indicating that YouMove’s effectiveness is not dependent on movement difficulty.

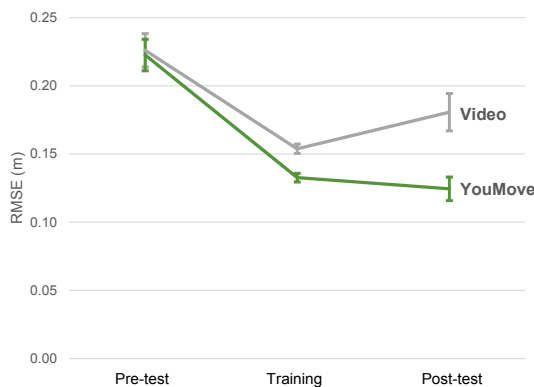


Figure 10: Pre-test, training, and post-test scores for the YouMove and Video conditions.

An ANOVA was also conducted on the post-test scores. The condition was significant ($F_{(1,7)} = 9.96, p = 0.02$) with scores of 0.12m for YouMove and 0.18m for Video, an improvement of 33% (Figure 10). The movement type was also significant ($F_{(1,7)} = 114.2, p < 0.01$), with scores of

0.10m for ballet movements and 0.19m for abstract movements, indicating that ballet movements were easier.

Stage Usage Analysis

For the *YouMove* condition, most participants performed the training guides in sequence, although some participants did choose to replay guides before continuing. After unlocking all stages, there was no clear preference, but users seemed less likely to return to the posture guide. This is likely due to the increased time the posture guide requires. The guide use by trial is visualized in Figure 11.

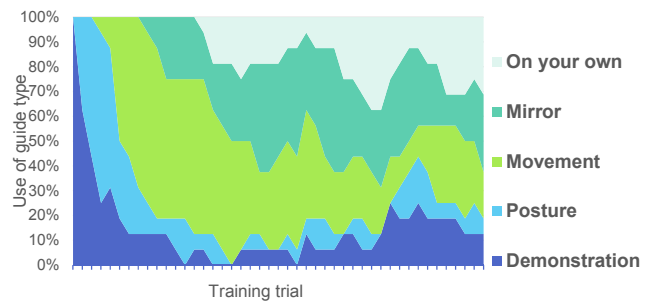


Figure 11: Frequency of use for each guide type for all participants during training with the YouMove system.

DISCUSSION

Overall, we were pleased to see that when comparing pre-test and post-test results, learning increased by more than a factor of 2 (44% vs. 20%) with the YouMove system.

Analysis of learning

Across all conditions, participants appeared to learn the movement within the first few training trials (Figure 9) and then reached a plateau, with relatively constant performance. In the case of the video condition, this was likely because the trainee had truly plateaued, and learned all of the usable information from the video. With the YouMove condition, however, the gradual reduction of guidance meant that the trainee was continually learning the movement and making up for the lack of guidance with increased skill, resulting in consistently high performance.

The retention scores show improved learning with YouMove, as performance is maintained on the retention tests. In contrast, retention scores for video are worse than in training, evidenced by the upward slope in Figure 10.

While training with the YouMove system took longer (i.e., average of 20 minutes compared to 11 minutes with video), this cannot fully explain the results of learning, as the number of exposures to the movement remained constant. Additionally, performance on the video condition plateaued after approximately 8 repetitions, and unlike the YouMove condition, the stimuli did not change and so it is unlikely that any additional learning would occur if the video condition had been repeated for 20 minutes.

Mirror interaction

Dynamic lighting allowed the system to shift the attention of the user, and did not seem to distract participants. The

dynamic lighting provided the intended effect and enabled the screen to be multiplexed, allowing the trainee to use both their reflection and the projected guide.

Some participants had difficulty lining up the reflection of their hand with the buttons for the reflection selection technique. To mitigate these problems, we have since added virtual cursors that appear if the system has been waiting more than five seconds for a selection to be made. These cursors provide novice users with the necessary feedback to correct their movement, yet do not interfere with the efficient reflection-based interaction.

Query by example

While the ability to search for similar movements was not tested in the user study, during informal use it proved to be fairly robust and useful. For novices, this feature will allow querying of large databases without having to know the name of the movement. This is especially useful in domains that have complex or cryptic names for their movements.

During training, the trainee's performance is recorded in the same format as the trainer's. This allows users to upload their movements for others to learn from, and enables them to review their prior performances at a later date.

Authoring

The authoring system proved to be easy to use. No formal study was conducted, but it was sent to an untrained user and they were able to record and annotate several movements with minimal effort. The movements were successfully transferred and used in our training system.

The authoring system benefits experts as it allows for content expansion. One minute of content creation generates approximately ten minutes of unique training content through the use of stages which vary the method of presentation and provide feedback. This allows experts with limited time to produce a useful quantity of content with minimal time investment.

Training stages

The Demonstration stage was well liked by users. This is likely due to familiarity with video as a teaching tool, and the ability for participants to easily understand the content. Some participants did, however, comment on the difficulty in judging limb placement and depth from video.

The Posture Guide allowed users to correct the movement, and many errors in positioning were corrected using the posture guide and depth callout. Though similar in nature to the posture guide, the movement guide was one of the most preferred guides. It provides practice guidance at full-speed while still providing useful feedback.

The Mirror Guide and 'On Your Own' guide were clearly valuable. Many participants felt that they had learned the movement, but when the guides were taken away they struggled to remember the sequence.

The Post-Stage Feedback was also well received. Five participants explicitly mentioned that the ability to move back and forth to change the viewpoint of the skeleton was a useful feature. The utility of the feedback is evident in the fact that many participants spent a substantial amount of time on this stage. They were aware that the amount of exposures to the movement was kept constant, and that any time spent on the feedback screen would lengthen the experiment, yet many participants still examined each keyframe.

LIMITATIONS AND FUTURE WORK

The evaluation captures the combined effects of many novel elements. More work is needed to analyze each component and its contribution to learning. Long-term learning and retention are also important areas to explore.

One limitation of the system is related to the skeletal tracking of the Kinect. In particular, the Kinect has difficulty tracking movements that cause large amounts of occlusions. One possible improvement would be to leverage multiple Kinects [5]. Improvements in sensing technology will open new domains of training. Playing musical instruments, surgery, many sports, and other motor-skill based domains could benefit from such a system.

The use of RMSE as a measure of learning is limiting, as real world movements may differ in positioning, but convey the same movements. More work is needed to accurately model and assess full body movement similarity.

Another limitation is the large mirror required for the training system. While the presented implementation uses a half-silvered mirror as a display, the software could also run as a traditional video-based augmented reality system, as in [35]. This would be more accessible to users, but does not provide the real-time feedback that the mirror does. It would be interesting to better understand any learning difference between a mirror and video based system on various devices (large screen, small screen, etc.).

The addition of social features and richer gamification could also greatly help YouMove. One can imagine online yoga, dance or martial arts classes, with competition from online peer groups, but more work is needed to achieve this.

CONCLUSION

We have presented *YouMove*, a novel full-body movement training system. Our work contributes an augmented reality mirror implementation, a content creation system that allows anyone to easily record full-body movements, and a training system that uses the recorded movements to instruct a trainee using a series of stages of increasing difficulty. We have also presented a user study that shows that YouMove results in better short-term retention scores than a video-based learning approach. As implemented, the system is immediately useful for a number domains, such as therapy, yoga, and many types of dance. Adapting the techniques presented here to a wider range of movements opens new avenues for future research.

REFERENCES

1. Anderson, F., Annett, M. & Bischof, WF. (2012). Tabletops in motion: the kinetics and kinematics of interactive surface physical therapy. *CHI EA*, 2351-2356.
2. Anderson, F., & Bischof, WF. (2013). Learning and Performance with Gesture Guides. *CHI* 1109-1118.
3. Bau, O. & MacKay, W., (2008). OctoPocus: a dynamic guide for learning gesture-based command sets. *UIST*, 37-46.
4. Bederson, BB. (2004). Interfaces for staying in the flow. *Ubiquity* 5(27).
5. Butler, D. A., Izadi, S., Hilliges, O., Molyneaux, D., Hodges, S., & Kim, D. (2012). Shake'n'sense: reducing interference for overlapping structured light depth cameras. *CHI* 1933-1936.
6. Dance Central. (2010). *Harmonix*, Massachusetts, USA.
7. Dearborn, K., & Ross, R. (2006). Dance learning and the mirror: comparison study of dance phrase learning with and without mirrors. *Dance Education*, 109-115.
8. Doctorow, C. (2011). Haunted Mansion Hitchhiking Ghosts go digital, play high-tech pranks on riders. <http://boingboing.net/2011/04/06/haunted-mansion-hitc.html>.
9. Freeman, D., Benko, H., Ringel Morris, M. & Wigdor, D. (2009). ShadowGuides: visualizations for in-situ learning of multi-touch and whole-hand gestures. *ITS*, 167-172.
10. Fujinami, K., Kawsar, F. and Nakajima, T. (2006). AwareMirror: A Personalized Display Using a Mirror. In *Proc. Pervasive Computing*, 316-332.
11. Henderson, SJ. & Feiner, SK. (2011). Augmented reality in the psychomotor phase of a procedural task. *ISMAR*, 191-200.
12. Holden, MK. (2005). Virtual environments for motor rehabilitation: review. *Cyberpsychology*, 8(3), 187-211.
13. Janelle, CM., Barba, DA., Frehlich, SG., Tennant, LK., & Cauraugh, JH. (1997). Maximizing performance feedback effectiveness through videotape replay and a self-controlled learning environment. *Research Quarterly for Exercise and Sport*, 68(4), 269-279.
14. Judkins, TN., Oleynikov, D. & Stergiou, N. (2005). Real-time augmented feedback benefits robotic laparoscopic training. *Stud. Health Tech.* 119, 243-248.
15. Krueger, MW. (1977). Responsive Environments. *AFIPS*, 423-433.
16. Lehmann, KS., Ritz, JP., Maass, H., Çakmak, HK., Kuehnappel, UG., Germer, CT., & Buhr, HJ. (2005). A prospective randomized study to test the transfer of basic psychomotor skills from virtual reality to physical reality in a comparable training setting. *Annals of surgery*, 241(3), 442.
17. Li, W. Grossman, T. & Fitzmaurice, G. (2012). GamiCAD: A gamified tutorial system for first time autocad users. *UIST*, 103-112.
18. Li, FCY, Dearman, D, & Truong, KN. (2009). Virtual shelves: interactions with orientation aware devices. *CHI*, 125-128.
19. Microsoft Research, (2012). Hololector. research.microsoft.com/apps/video/dl.aspx?id=159487.
20. Ng, A., Lepinski, J., Wigdor, D., Sanders, S., & Dietz, P. (2012). Designing for low-latency direct-touch input. *UIST*, 453-464.
21. Palmiter, S., & Elkerton, J. (1991). An evaluation of animated demonstrations of learning computer-based tasks. *CHI*, 257-263.
22. Pierce, JS. Forsberg, A., Conway, M, Hong, S. & Zeleznik, R. (1997). Image Plane Interaction Techniques in 3D Immersive Environments. *IBD*, 39-43.
23. Pongnumkul, S., Dontcheva, M., Li, W., Wang, J., Bourdev, L., Avidan, S., & Cohen, MF. (2011). Pause-and-play: automatically linking screencast video tutorials with applications. *UIST*. 135-144.
24. Proteau, L., Marteniuk, R.G., & Lévesque, L. (1992). A sensorimotor basis for motor learning: Evidence indicating specificity of practice. *Exp. Psych.*, 44(3), 557-575.
25. Sato, H., Kitahara, I., & Ohta, Y. (2009). MR-mirror: a complex of real and virtual mirrors. *VMR*, 482-491.
26. Schmidt, RA. (1991). Frequent augmented feedback can degrade learning: Evidence and interpretations. *Tutorials in motor neuroscience*. 59-75.
27. Schmidt, R. A., & Wulf, G. (1997). Continuous concurrent feedback degrades skill learning: Implications for training and simulation. *Human Factors*, 39(4), 509-525.
28. Shea, C. H., Wulf, G., Park, JH., & Gaunt, B. (2001). Effects of an auditory model on the learning of relative and absolute timing. *Motor behavior*, 33(2), 127-138.
29. Shoemaker, G., Tang, A., & Booth, K. S. (2007). Shadow Reaching: A New Perspective on Interaction for Large Wall Displays. *UIST*, 53-56.
30. Sigrist, R., Rauter, G., Riener, R. & Wolf, P. (2012). Augmented visual, auditory, haptic and multimodal feedback in motor learning: A review. *Psych. Bull.Rev.*, 1-13.
31. Sodhi, R., Benko, H. & Wilson, A. (2012). LightGuide: Projected visualizations for hand movement guidance. *CHI*, 179-188.
32. Todorov, E., Shadmehr, R. & Bizzi, E. (1997). Augmented feedback presented in a virtual environment accelerates learning a difficult motor task. *Motor Behavior* 29(2) 147-158.
33. Ushida, K., Tanaka, Y., Naemura, T., & Harashima, H. (2002). i-mirror: An Interaction / Information Environment Based on a Mirror Metaphor Aiming to Install into Our Life Space. *Artificial Reality and Telexistence*, 113-118.
34. Van Merriënboer, J. J., & Sweller, J. (2005). Cognitive load theory and complex learning: Recent developments and future directions. *Educational psychology review*, 17(2), 147-177.
35. Velloso, E., Bulling, A. & Gellerson, H. (2013). MotionMA: Modelling and Analysis by Demonstration. *CHI* 1309-1318.
36. Vogel, D. & Balakrishnan, R. (2005). Distant freehand pointing and clicking on very large, high resolution displays. *UIST*, 33-42.
37. Vogel, D. & Balakrishnan, R. (2004). Interactive public ambient displays; transitioning from implicit to explicit, public to personal, interaction with multiple users. *UIST*, 137-146.
38. Ward, P., Hodges, NJ., Williams, AM., & Starkes, JL. (2004). Deliberate practice and expert performance. *Skill acquisition in sport: Research, theory and practice*, 231.
39. Winstein, CJ., & Schmidt, RA. (1990). Reduced frequency of knowledge of results enhances motor skill learning. *Exp. Psych.: Learning, Memory, and Cognition*, 16(4). 677-691.
40. Wollersheim, D., Merkes, M., Shields, N., Liamputtong, P., Wallis, L., Reynolds, F., & Koh, L. (2010). Physical and psychosocial effects of Wii video game use among older women. *International Journal of Emerging Technologies and Society*, 8(2), 85-98.
41. Wulf, G. Raupach, M. & Pfeiffer, F. (2005). Self-controlled observational practice enhances learning. *Research Quarterly for Exercise & Sport*, 76, 107-111.
42. Wulf, G., Shea, C. & Lewthwaite, R. (2010). Motor skill learning and performance: a review of influential factors. *Medical Education*, 44, 75-84.
43. Yang, U. & Kim, GJ. (2002). Implementation and evaluation of just follow me: An immersive, VR-based, motion-training system. *Presence* 11(3) 304-323.