Limb-O: Real-Time Comparison and Visualization of Lower Limb Motions

Michael Gammon
University of Manitoba
michael.gammon@umanitoba.ca

Ritesh Udhani
University of Manitoba
udhanir@myumanitoba.ca

Parth Patekar
University of Manitoba
parth.15je001620@am.ism.ac.in

Yumiko Sakamoto
University of Manitoba
umsakamo@umanitoba.ca

Carlo Menon
Simon Fraser University
cmenon@sfu.ca

Pourang Irani
University of Manitoba
pourang.irani@cs.umanitoba.ca

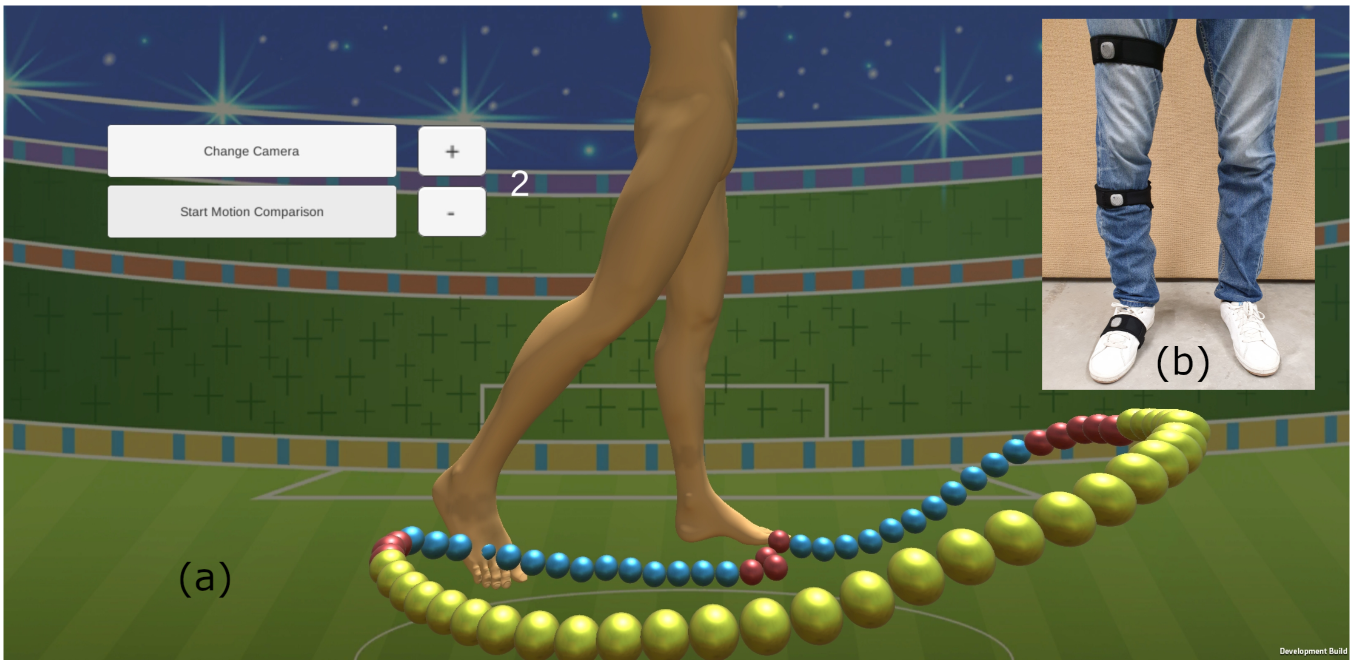


Figure 1: (a) Limb-O is a virtual training application that uses orbs placed in front of the user's leg to help guide their motion, for applications such as training or rehabilitation. (b) The user wears three IMU sensors along their leg that track their motion.

ABSTRACT

Since the rise in popularity of video-sharing platforms such as Youtube, learning new skills from the comfort of one's own home has become more accessible than ever. Though such independent learning methods are useful, they lack the real-time feedback component of being in the same room with an expert, which is why expensive private coaching sessions remain desirable. Accordingly, we propose Limb-O (orbs for limb movement visualization), a real-time quantitative virtual coach application for learning lower-limb motions through motion comparison. The proposed application turns the practice of things like sports motions into a game that highlights imperfections and allows for tracking of progress over time. A user validation study was run which confirmed that Limb-O outperforms traditional video learning methods both quantitatively and qualitatively, by providing objective feedback that keeps users engaged.

KEYWORDS

Lower Limb, Motion Comparison, Virtual Coach, Sports Training, Temporal Alignment, Rehabilitation, Telerehabilitation

ACM Reference Format:
Michael Gammon, Ritesh Udhani, Parth Patekar, Yumiko Sakamoto, Carlo Menon, and Pourang Irani. 2020. Limb-O: Real-Time Comparison and Visualization of Lower Limb Motions. In *11th Augmented Human International Conference (AH '20), May 27–29, 2020, Winnipeg, MB, Canada.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3396339.3396360

AH '20, May 27–29, 2020, Winnipeg, MB, Canada

ACM ISBN 978-1-4503-7728-7/20/05…$15.00
https://doi.org/10.1145/3396339.3396360

1 INTRODUCTION

Online tutorials, either text, picture, or video-based, have made the barrier to entry to many traditionally difficult fields lower than ever. Picking up a new skill - whether it be learning a musical instrument, advanced sports techniques, or beginner aerobic exercise routines - is made simpler with such references, while avoiding the high monetary cost of a private teacher. While this is clearly positive, most methods of remote or independent learning are not yet comparable to working side-by-side with an expert in the field when learning low-level concepts [6]. This is likely due to difficulty quantifying the similarity between the beginner's attempt and the source that they are receiving the information from, which could lead to skill acquisition taking longer than necessary.

To further accelerate low-cost & accessible training, advances in virtual learning and coaching technologies are desirable. Ideally such technologies should not only provide a reference to learn from, but also a feedback method for automatically quantifying the difference between a user's performance and an expert's in order to make suggestions on how they can further improve. It has been shown that VR training that highlights such differences can perform better than training from a human expert [16]. With this motivation in mind, we introduce Limb-O - a low cost, portable virtual coach for lower-limb motion comparison. Limb-O in its current form is presented as a mobile Android smartphone application that pairs with three Bluetooth IMU sensors capturing the motion of the thigh, shank, and foot. An "expert" can capture lower-limb motions in the app (e.g., a particular karate kick or a leg motion for a dance move) that a user will attempt to replicate. The visualization, presented as orbs the user will need to follow with the foot, will highlight the similarities and differences between the user's attempt and the expert's motion. Limb-O also provides users with a score after each attempt. This helps not only with confidence while learning, but also as a motivator [4].

Applications for comparing motion (or other forms of temporal media) do currently exist (InterPoser [13], Super Mirror [9], betaCube [18], TIKL [8]), but typically are either strictly qualitative in nature or only work for the comparison of static poses. We propose Limb-O as a solution for the quantitative comparison of dynamic lower-limb motions. Our user study reveals that practicing motions with Limb-O leads to a better outcome than the video comparison method does. Further, in the users' subjective experiences, Limb-O was better for understanding the task, better for focus, and all-around preferred over video comparison.

2 RELATED WORK

2.1 Spatio-Temporal Media Comparison

The main challenge in creating a virtual coach able to compare two sets of lower limb motions in a meaningful way stems from the fact that temporal media comparison is, in general, a complex task. Tharatipyakul [14] suggests that visual temporal media (such as animated joint angle data, in our case) can be uniquely difficult to compare since the relationship with respect to time is usually equally as important as the relationship between the two (or more) media instances. She argues that visual design for such comparison is "almost non-existent". In most cases people rely on memory for such tasks, which usually results in the loss of useful information.

One approach that is often used in video comparison is that of juxtaposition. The problem is that when the user is watching one media item, they are missing the content from the other media item as time progresses - "*the temporal nature of the media necessitates additional or different design elements to support the comparison*" [15]. Other possibilities for comparison include superposition and relationship encoding [15], which are typically thought of as better options, although present new challenges.

The playback method for the temporal media can either be sequential or in parallel. Playing the media in parallel is typically a superior method for comparison (due to a high cognitive load not hindering the effectiveness of the comparison) [15], but brings up the additional problem of content synchronization/temporal alignment. This means that the portions of the media being played back simultaneously must share some similarities that are conducive to comparison which can be challenging to facilitate as the two pieces of media are not always of the same (or even similar) lengths. Studies for automating [17] or crowd-sourcing [12] content synchronization exist, though it is not always clear what content synchronization means when the structure/duration of the media are different. Should one piece of media be the "master", while the other "slave" media jumps to different parts/varies playback speed? What about a real-time comparison to a piece of pre-recorded media? These questions do not have a general answer, and the answers are often unclear - more studies and justifications for one method over another are needed [15].

Tharatipyakul argues that one proposed solution to the above problem is to remove the temporal component altogether, effectively turning a video (or other pieces of temporal media) into a series of keyframes [15]. Comparison between a set number of keyframes selected from each piece of media drastically reduces complexity, though there is a trade-off with the amount of information available. We build off such an idea in our own application by reducing the complexity of the temporal component of comparison and concern ourselves mainly with the spatial component - our interest in the temporal component is primarily with the order in which events take place rather than the actual time between events.

2.2 Methods For Visualizing Differences In Videos

Another issue once the pieces of the media have been aligned is how one measures and visualizes dissimilarities over time. For the case of comparing two videos, Balakrishnan et al. [2] developed an algorithm for measuring dissimilarity between two videos at each point in time, and using this measure of similarity at each point in the frame they visualize outlines where the two videos diverge from one another. Such a visualization method is simple, yet "dramatically enhances one's ability to notice differences" [2]. This is because "humans are better at interpreting spatio-temporal matches than current computer vision algorithms, so instead of automatically matching video pixels, [they] choose to direct users to important areas of the videos and let them interpret the differences" [2].

Another method of interest for visualizing spatio-temporal differences is a "ghosting" method. Ghosting is "a technique that uses machine learning to compute optimal player trajectories and play outcomes" [10] for sports training. One could imagine a semi-transparent version of the ideal motion being displayed alongside the user's true motion, similar to a "ghost car" in racing video games representing the player's best time.

2.3 Visualizing Comparison of Human Limb Motion and Positioning

Comparing human limb motion as opposed to motion in general is a unique and interesting problem. For instance, the leg is made up of three independent rotation points (hip/knee/ankle), and for two legs to be considered to be occupying the same orientation, the pitch/roll/yaw of all three of these joints must match (within some threshold) at that instant in time - nine parameters per leg. Such a large amount of information makes visualization challenging.

Marquardt et al. [9] visualized the latter scenario through juxtaposition of an animated stick figure with real-time skeleton tracking data from a Kinect-based system for evaluating the accuracy of ballet poses. When the pose for a given limb was correct (up to a certain inaccuracy threshold set by the researchers), it would change colour on screen. The real-time joint angles, relative to the target joint angles, were also displayed on-screen via moving wheels. For static poses this visualization method is quite intuitive, although it does not extend well to dynamic motions. In future iterations of the system, Marquardt et al. plan to define each step in a dynamic motion as a series of static poses to iterate through. Unfortunately, this plan is not generalizable to a smooth dynamic motion such as a soccer kick - the smooth temporal component of a kick makes it challenging to observe the desired motion while precisely following it at the same time.

An approach for dynamic motion comparison for rock climbers is presented by Wiehr et al. with betaCube [18]. The apparatus enables many different kinds of movement visualizations. One such visualization involves the recording and re-projection of an expert climber's run back onto the wall (that could either be watched from the ground, raced against, or mimicked in real-time [referred to as "shadow climbing"]) which gives detailed, repeatable information about the ground truth run but does not directly draw comparisons to the user's movements. Alternatively, the apparatus can also track which holds were used in succession by the expert, and highlight them via the projection of markers around the next holds in the climb. This ensures that the user follows the ground truth path closely, but does not give feedback on the actual nuanced joint movement between holds which is key to becoming a strong climber.

Aside from visual feedback, joint angle information has also been compared using tactile feedback. Lieberman et al. [8] developed a vibrotactile feedback suit as a means of providing real-time feedback. They propose that tactile feedback directly engages our motor learning systems as opposed to something like auditory feedback, which is abstract and requires the creation of a mental model. The suit works by delivering vibrations proportional to the amount of error along that axis, effectively nudging the user's motion back on track. This method is interesting, although it requires the use of an expensive Vicon tracker and feedback suit.

2.4 Tracking Methods for Motion Comparison

In previous work, tracking motions for comparison typically depends on capturing position in either two (x,y) or three (x,y,z) dimensions and drawing direct comparisons between the positions of specific features in a local Cartesian coordinate system. This is seen, for instance, in [2] and [13] where a standard RGB video is captured and the pixel locations of certain features are compared. Such comparisons have also been made using Vicon or CAVE-based systems [8] for more accurate tracking results. These systems have two inherent disadvantages - first, the capture methods are inconvenient and non-portable. In the case of video comparison, the backgrounds of both videos should be identical to avoid picking up false positives. This means that once a capture location has been chosen, all motions to be compared must take place there without any relocation of the camera. In the case of the Vicon or CAVE (or a camera with a green screen), the captures must take place in an unnatural laboratory setting and cost may be prohibitive. Additionally, the use of Cartesian position as a means of comparison makes it difficult to compare the motions of people with different heights and weights without laborious editing to the raw capture files, as a similarly performed motion from, for instance, both an adult and child will lead to very different positions of the limbs in Cartesian space.

Some of the aforementioned problems were solved in [9] where a Kinect camera captures the skeleton, and thus the joint angles of the user, in order to compare their poses to a ground truth set of poses. This method of using joint angles as opposed to position makes for a much more convenient comparison between people of different shapes/sizes and is not dependent on the background of the camera capture. That being said, it is still necessary for the user to stay within the field of view of the Kinect. This makes such a solution impractical for anything gait-related, or anything that is more naturally performed outdoors/in public.

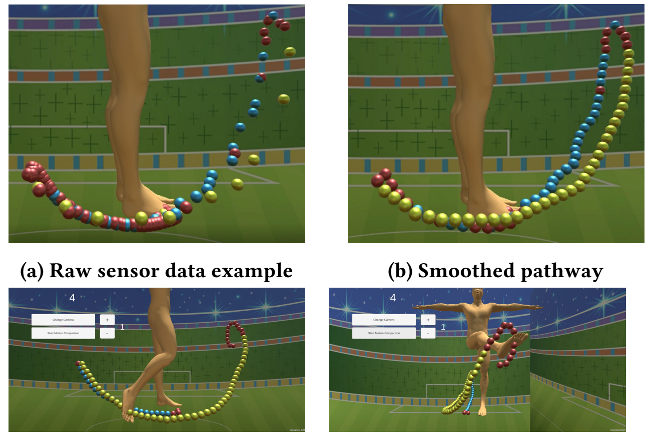
2.5 Increasing Motivation for Home Rehabilitation

In order to increase motivation and perseverance for tedious tasks, often designers use gamification methods [4]. This was done for rehabilitation tasks in [11] by turning the therapy-related exercises into controls for a game using accelerometers, gyroscopes, and flex sensors. The platform focuses on acting as a personal trainer for the patient that also collects the patient's performance data during the training for later evaluation. Such data collection is useful for tracking progress over time when the platform is used on a regular basis. Even without gamification, [7] found that simply having an app with remote support was enough to increase adherence to exercise programs when compared to paper handouts.

3 LIMB-O

We introduce Limb-O - a novel solution for the real-time comparison and coaching of lower limb motion in humans. Limb-O in its current form is presented as an Android application designed to act like a virtual coach for sports activities primarily involving the lower body (such as soccer, martial arts kicks, dance moves, and repetitive leg rehabilitation exercises). With these applications in mind, we aim to find a way to quantify the similarity between a user's attempt at a given motion to a ground truth input for that same motion (ideally from an expert) in real time while also providing feedback as to where improvement is needed.

The application works by breaking up a ground truth lower-limb motion into a series of keyframes represented by collectable orb objects, which are sequentially collected by the user's virtual foot to increase their score. A perfect replication of the reference motion will result in the destruction of all on-screen orbs, and a perfect score of 100. Otherwise, the remaining orbs give feedback to the user as to where improvement may be necessary.



(a) Raw sensor data example **(b) Smoothed pathway**

(c) Soccer Kick Side View **(d) Soccer Kick Front View**

Figure 2: Limb-O application

3.1 Hardware and Data Stream

Our basis for lower limb comparison is entirely orientation-based. The complete configuration of a leg relative to the pelvis can be described through nine parameters - namely the orientation of each leg segment (thigh, shank, foot). Each of these segments require three rotation angles for a full description (although, the knee is regarded as a hinge joint which to a good approximation is only allowed flexion/extension [3]). In order to obtain this orientation data in real-time while remaining lightweight, portable, and affordable, we used the MetaMotionC (MMC) sensors from Mbientlab (which includes an on-board sensor fusion algorithm) to stream quaternions from each leg segment in real-time. For the purpose of the user study described in this paper we only compared the motions of one leg at a time, meaning a total of three sensors were strapped on each participant and streaming data (although the application can be extended for up to seven sensors).

The portability/convenience of these sensors (which can easily stream to a standard smartphone over Bluetooth) along with the affordable price makes them ideal for use outside of the lab (i.e. practicing a soccer kick with a ball in a field), and makes them an attractive option for in-home rehab applications. It is much more affordable for a patient with severe mobility issues to buy sensors to pair with their smartphone than to pay a physiotherapist for in-home visits. In future work, we look into ways to make this process entirely remote - the therapist could upload new motions to the cloud, and track the patient's progress without ever being in the same room.

Since the MMC sensors are not able to stream data directly to Unity, an intermediate Android application was written to facilitate the connection of the boards and subsequently stream the data continuously in the background. This data was then broadcast to a Unity 3D application which received the quaternions from each board and used them to update the orientation of the leg joints (hip/knee/ankle) on a 3D humanoid model every frame by calculating the difference in orientation from a calibration T-pose. The joint angles were validated using a Vicon motion capture system. If the data streaming rate (in our case, $\sim 90Hz$) and frame rate are high enough, this setup results in the appearance of a smooth lower body motion on the 3D model that mirrors that of the user.

3.2 Ground Truth Data Collection and Path Interpolation

In order to compare two sets of motions, it is first necessary to collect a ground truth data set to serve as the basis for comparison. To do this, an "expert" user wears the three sensors on their leg and performs a desired motion with the "capture" mode running on the app. While this happens, the (x,y,z) positions of the 3D model's toe are written to a file in every frame. This position is a direct function of the joint angles of the hip/knee/ankle and is purposely independent of the position of the real user's foot or leg size.

When the position data is written to file, it is messy in general. Typically after pressing the button to begin logging, the expert's foot will sit static at the starting position for a while before the movement begins. This leads to a surplus of data points all collected around the same position. Then when the motion is being performed, the velocity of the foot varies, leading to uneven spacing between the logged points. For these reasons it is necessary to apply a runtime filtering technique to evenly space data points along the desired path. This way, when a user tries to compare their own motion to the ground truth, a percentage of matching data points will actually serve as a meaningful scoring system.

We use a direction vector based approach to smooth the pathway generated by the data points from the sensors. With this approach, a linear interpolation of the pathway between the data points was generated. Figures 2a and 2b show the before/after pathways.

3.3 Visualization and Scoring Method

When a user is ready to start learning a lower limb motion using Limb-O, they load the data set described in the previous section through an on-screen menu and the interpolated data set is rendered as a set of coloured collectable orb objects. The goal of the virtual coach game is to collect these orbs in the order they were rendered by passing the toes of the 3D model through them. Recall that since our system is entirely orientation-based, this is independent of the users' leg size and thus one ground truth data set can be used for comparison by anyone. When a user successfully collects an orb, it will disappear from the screen, and the score is incremented. After the motion is completed, the user can see where they require improvement by observing where large amounts of orbs are left over.

When these orbs are rendered on screen, a simple linear velocity calculation is done by comparing the positions and timestamps of neighbouring data points. If the approximated linear velocity is less than some threshold (in our case 5×10^{-4} Unity distance units/ms), then the orbs are rendered in a red colour. These orbs are directionally ambiguous and can be collected by passing the model's toes through in any direction. If the approximated linear velocity is greater than the velocity threshold, then they are either coloured gold (forward motion of the toe) or blue (backwards motion). These orbs can only be collected if the direction of motion is appropriate. The reason for this is that some motions (such as some soccer kicks) include both a forward and backward component that cross paths, and it is not desirable for the user to accidentally start collecting orbs from the wrong part of the motion. Currently this is only implemented for the forward/backward direction, but future iterations could easily implement a switch to apply this colour coding to left/right instead, depending on the desired motion. We take further measures to avoid accidental collection by having a window for collection relative to the last collected orb. For our study this window size was 26, meaning that it is impossible to collect an orb more than 26 positions ahead of that which was last collected. This prevents the accidental collection of orbs from later in the motion but also introduces the problem where if a large chunk of orbs are missed, then collection becomes impossible for the rest of the motion. The window size for our study was chosen through trial and error, in an effort to avoid the problem where orb collection is cut off as often as possible. The ideal window size will depend on the motion itself and the number of orbs concentrated in small volumes.

Since the temporal component of the media is broken down into static keyframes (orbs), the only time-dependence is the order in which the orbs are collected, along with the direction of motion (forwards/backwards) of the 3D model's foot. This greatly reduces complexity (as was suggested in [15]) while still providing useful feedback to the user. Since these orbs exist in 3D space, we included a "Change Camera" button in the app that will switch the user's view between front (Figure 2c) and side (Figure 2d). This can be done at any time, and the user may perform the orb collection in whichever view is more convenient for the task at hand.

4 VALIDATION STUDY

In our study, participants were asked to replicate two lower limb movements which had been performed and recorded in advance by the researchers. The motion replication was completed by providing real-time feedback through Limb-O in the experimental condition (presented on a Samsung Galaxy S8 5.8" smartphone), as well as using video feedback in the control condition. The experiment aimed to see whether the participants' performance varied depending on the technologies they used (i.e., Limb-O vs. typical video learning) in motion replication, both quantitatively and qualitatively (i.e., in the user's experience). As this user study is simply meant to serve as an initial proof of concept for motion replication, rather than having real sports experts input their data as a point

(a) Side View **(b) Front View**

Figure 3: Video livestream views during "soccer kick" task.

of comparison, the authors instead input arbitrary lower limb motions themselves that would be simple enough for an average participant to achieve a decent score on.

We aimed to make the video control scenario as comparable to Limb-O as possible; when the ground truth data sets for Limb-O were collected, we recorded the "expert" (i.e., reference video) from two different angles (to be comparable to the two viewing angles available in the app) and allowed the participant to view these videos before each attempt at motion replication. In order to provide an equivalent feedback to the one provided by Limb-O, we live-streamed the participants' motions to a smartphone during their trials so that they could watch their motions from both the side (Figure 3a) and front (Figure 3b) angles as they happened, analogous to watching the 3D model move in Limb-O without the feedback of the collectable orbs.

While participants were completing their trials using the video comparison method, we had Limb-O running in the background (with the visualization hidden from the user). This was done so that orbs would still be collected to quantify how close they were to performing the ground truth motion, giving us a number to compare to their trials with the experimental condition (Limb-O).

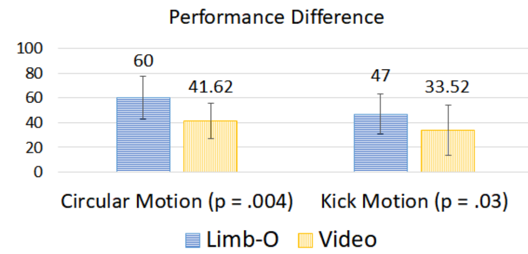
4.1 Experimental Method

A within design study was conducted where sixteen (16) local university students (F = 4) participated to explore the efficacy and the user experience of our visualization approach in Limb-O. The participants were asked to test a new piece of technology that would act like a virtual coach, giving real-time feedback as they try to replicate two different lower limb motions - a soccer kick, and a circular leg sweep motion.

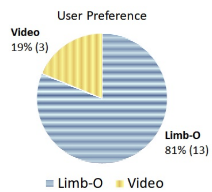
First, participants were greeted and asked to turn off and remove any smart devices, and the the steps in the study were explained. They were then asked to sign a consent form and fill out a general demographic questionnaire. Next, three MMC IMU sensors were strapped onto the participant's right leg (thigh/shank/foot - see Figure 1b). There were two tasks (soccer kick & circular leg sweep) to complete with each of the feedback systems, with the second task serving as a distraction to minimize the learning effect (e.g., 1st: Soccer Kick with App; 2nd: Circular Sweep with Video; 3rd: Soccer Kick with Video; 4th: Circular Sweep with App). These tasks were also selected to explore how the app's performance differs between simple and difficult motions - the soccer kick task was pointedly more challenging for participants as it required a much larger range-of-motion than the circular leg sweep. The tasks were coordinated in four different orders to minimize the order effect.

The first time a task was completed using Limb-O, we had the participant stand on a line of blue tape on the floor with feet shoulder-width apart while the sensor calibration was completed (Figure 1b). We demonstrated the app to the participant, and illustrated that if they moved their right leg while keeping their left foot planted on the ground, the 3D model on screen would move along with them. We also demonstrated that the viewing angle could be changed by pressing the "Change Camera" button. We then showed the participant how to load the task that they were to complete, and had them watch the animation of the 3D model completing the desired task through the rendered orbs. Participants were told that their goal was to collect as many of the orbs as possible by moving the 3D model's (i.e., their avatar's) right toes through these orbs as they mimicked the motion seen in the animation. Participants were told that collecting the orbs in succession would cause the score to increase, with a maximum score of 100 if all orbs were collected. We also explained the relationship between the direction of motion and orb colour to the participants (see Section 3.3 for more detail). The participants were then instructed to practice the task using Limb-O until they were confident in their performance, at which point they performed three final "recorded" trials. Although we did not quantitatively record motion speed, we asked participants to perform the motions smoothly as opposed to slowly collecting each orb (which would defeat the purpose of the app). The importance of using both camera angles to learn the motions was also emphasized. For the recorded trials participants were urged to perform the task to the best of their ability, and the three scores were noted by the researcher. In order for the sensors to stay appropriately calibrated with the user's orientation, we ensured that they returned to the same pose on the blue tape line after each trial. We then asked the participants to fill out the portion of our questionnaire regarding the relevant task paired with the Limb-O app.

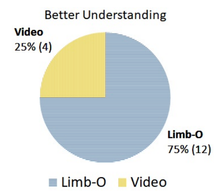
The first time a task was completed using the video feedback method, we had the participant stand on the blue tape line noted above. From there we explained how to navigate to the pre-recorded reference motions for the task that they were meant to emulate, and also showed them how to access the live streams of their own movement. The participants were asked to switch between these reference videos and the live streams of their movements while practicing the task until they felt confident that they were emulating it as closely as possible. When they indicated that they were ready for the recorded trials, Limb-O was loaded on a separate phone by the researcher, and the sensors were calibrated. We again stressed that they should try to be as accurate as possible in the following trials. The researcher started the comparison mode in Limb-O without telling the participant, and then allowed them to start the recorded trials while using the video feedback method. The scores in Limb-O were noted by the researcher, and the participants were instructed to return to the calibration position between each trial. We then asked the participants to fill out the portion of our questionnaire regarding the relevant task paired with the video feedback method.



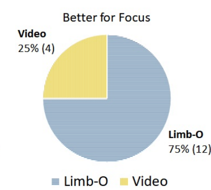
(a) User performance better with Limb-O



(b) Limb-O is preferred



(c) Limb-O provides better understanding



(d) Limb-O improves focus

Figure 4: Results.

Finally, the participant was asked to fill out a questionnaire regarding the comparison of the two feedback systems, and to complete the Ishihara colour blindness test.

4.2 Results

Note, throughout the analyses, assumptions for each analytical method were checked and appropriate methods were selected accordingly. Further, Cohen's rule-of-thumb [5] was applied in interpreting our effect sizes.

Participants 16 students from a local university participated in the study (F = 4) and their age ranged between 18 and 33 (M = 25.67, SD = 3.77). One participant skipped the age question. Two questions asked about their tech use experiences (e.g., "Generally speaking, compared to my friends and family, I am good at using most tech devices.") with a 7-point Likert scale. Due to high correlations between these items (p =0.01, r = 0.66), these scores were aggregated to create an index. Overall, our participants had high technology use experiences (M = 5.91, SD = 1.10).

Overall Performance (Limb-O vs. Video Comparison) Repeated ANOVAs investigated whether participants performance varied based on the technology they used. First, kick motion was investigated by comparing the *scores* between Limb-O vs. video comparison. Participants scored higher in the kick motion task when they used Limb-O (M = 47.00, SD = 16.24) compared to the time when they used the video comparison (M = 33.52, SD = 20.25), F (1, 14) = 5.81, p = 0.03, with a large effect (η_p^2 = 0.29). The same pattern emerged for the circular motion; Limb-O (M = 60.00, SD = 17.26) and video comparison (M = 41.62, SD = 14.31), F (1, 14) = 11.62, p = 0.01, with another large effect (η_p^2 = 0.45). (Figure 4)

Direct Comparisons Chi square tests explored the participants' preferences on the technology (i.e., Limb-O vs. video comparison). Four questions were asked (1: Which system gives a better understanding of leg movement did you feel?, 2: Which system is easier to learn the leg movement from?, 3: Now you have tried two systems. Which one do you prefer?, 4: Which system allowed you to focus on your task better?). In comparison to the video method, participants found the Limb-O visualization to be easier for understanding the leg movement, χ^2 (1, 16) = 4.00, p = 0.05, v = 0.25, allowed them to focus more on their task, χ^2 (1, 16) = 4.00, p = 0.05, v = 0.25, and simply preferred Limb-O, χ^2 (1, 16) = 6.25, p = 0.01, v = 0.39. However, there was no difference in the learning difficulty level (χ^2 (1, 16) = 0.25, p = 0.62), v = 0.02.

Kick movement Participants' perceived physical awkwardness in using Limb-O was assessed with a question: ("How physically awkward did you feel when you were using this method to replicate the leg movements?"). A repeated ANOVA revealed higher levels of awkwardness when participants used Limb-O (M = 3.88, SD = 1.02) in comparison to the video method (M = 2.75, SD = 1.24), F (1, 15) = 6.94, p = 0.02, with a large effect (η_p^2 = 0.32). Next, to explore the level of difficulty to copy the kick motion, a question was asked ("How easy was it for you to copy the leg motion?"). Overall, participants felt it was easier to copy leg movement using video comparison (M = 4.25, SD = 1.34) than with Limb-O (M = 3.06, SD = 1.48), F (1, 15) = 10.43, p = 0.01, (η_p^2 = 0.41).

Circular limb movement Parallel to the Kick movement, we explored participants' experiences of physical awkwardness and difficulty in copying the leg movement. Results from a repeated ANOVA indicated there was no difference in physical awkwardness across conditions (Limb-O: M = 3.13, SD = 1.20, video comparison: M = 3.31, SD = 1.49), F (1, 15) = 0.19, p = 0.67, but with a large effect (η_p^2 = 0.22). Regarding the perceived difficulty level in the copying of motion, using video comparison was perceived as easier (Limb-O: M = 3.75, SD = 1.39, video comparison: M = 4.75, SD = 1.48), F (1, 15) = 4.62, p = 0.05, with a large effect, (η_p^2 = 0.24).

Across movements First, "participants' interest in using the technology in the future" was assessed with three questions (e.g., "How interested would you be in using this technology if you were to practice some leg movements for playing sports?"), and the responses to these three questions were highly correlated with each other (ps< 0.015, rs > 0.60). Thus, an index was developed to indicate the participants' general interest in using the technology for learning Lower Limb movement. No cross-technology effect was found (p = 0.34). Next, two questions explored participants' general positive feelings associated with the technologies (e.g., "How much did you enjoy using this technology?"). Again, due to high correlation, (rs> 0.86, ps < 0.00001) we created an index for participants' pleasant feelings associated with their technology use. A significant technology type effect was found (Z = -2.08, p = 0.04, with a large effect r = 0.52; Limb-O: M =5.34, SD = 1.35, video comparison: M = 4.31, SD = 1.55). Participants reported more pleasant feelings when they used Limb-O over video comparison. Finally, we explored whether the participants' interest in using each technology for the future can be predicted by their pleasant feelings and their actual performance score. To explore this question concerning Limb-O, a multiple regression analysis was conducted with enter method: Interest in future use as a DV and Participants Leg Movement Performance Score and their Interest in Using the Limb-O as predictors. Note we report R^2_{adj} due to our smaller N size for regression analysis. R^2_{adj} = 0.328, indicating the model explained approximately 33% of variability in the participants' interest in using Limb-O; F(1,14) = 4.41, p = 0.037. While participants' pleasant feelings contributed to the model significantly, b = 0.649, p = 0.025, their actual score did not, b = 0.004, p = 0.99. Finally, the same model was tested for video comparison. R^2_{adj} was 0.78; the model explained approximately 80 % of variability in the participants' interest in using the video method, and it was significant F(1,14) = 13.65, p < 0.001. While participants' pleasant feelings contributed to the model significantly again, b = 0.826, p < 0.001, their actual score did not, b = -0.227, p = 0.102.

5 DISCUSSION

5.1 Discussion of Results

Overall, our study confirmed Limb-O's stronger efficacy as opposed to the video comparison method. This shows that Limb-O has great potential as a practice method, as the user is more likely to be practicing the motion similarly to how the ground truth was inputted. One issue with the qualitative comparison is that even with the prompt to perform the motions with Limb-O quickly and smoothly, a few participants got distracted by the orb collection aspect and attempted to slowly collect each object. Such distractions might be a shortcoming of our motion comparison method.

Users noted that perceived difficulty was higher when using Limb-O than when using the video comparison method for both of the tasks. Besides a lack of familiarity with the new technology, this is likely due to the fact that Limb-O is less forgiving of deviations from the ground truth data set, whereas with the video method, users wouldn't receive any feedback on such deviations. This leads to the feeling of perceived difficulty replicating the motion; while it may seem somewhat counter-intuitive, this perceived difficulty might have been a key to our positive outcome. In fact, for both tasks, the actual scores were higher when using Limb-O. The perceived challenge will lead to a user being more careful and deliberate with their motions, and learning the motion more accurately in the long run. In the open-ended questionnaire, many reported liking the fact that they could see concrete evidence of where they performed poorly, making the motions easier to learn even if actually performing the motion was more challenging.

Similarly, for the more challenging task (the soccer kick), users reported feeling more physical awkwardness using Limb-O than with the video comparison method. For the less challenging task (circular motion) this difference was not significant. This could be attributed to a similar reason as described above. Since the app is less forgiving, it is more likely to push a user to try harder to truly replicate the motion, even to the point of feeling slightly physically uncomfortable if that is what proper replication of the motion requires.

Users also reported more pleasant feelings when using Limb-O over the video comparison method. We found that the participants' "interest in using this technology in the future" could be predicted based on these positive feelings. Since repeated use of such a technology is necessary for developing expertise, this bodes well for the gamification aspect built in to Limb-O. This could also imply an increased motivation for rehabilitation applications where often the tasks to complete are challenging and uncomfortable.

5.2 Limitations and Future Directions

One of our main concerns with the current implementation of Limb-O stems from the typical performance gap between an expert and a beginner. Performing a karate kick even remotely similar to a black belt may be unrealistic for the average user, and thus the feedback provided by the lack of orbs collected might have only little use in such cases. In a future versions we intend to implement motion interpolation between a user's best attempt and the expert's motion, based on the work done in [13] with InterPoser. This way we could provide an intermediate motion path for the beginner to replicate that is closer to the expert's motion while still being a realistic goal.

In future iterations of this app, we would also like to explore being more concerned with the minutia of the movements - namely introducing a further velocity dependence, and also being more precise with orientation. In its current form, since orbs are only placed over the path that the model's toe follows, there are multiple states that the leg could occupy to collect a given orb. Adding additional orbs to collect on the knee joint (and potentially the ankle joint) removes the ambiguity, although could lead to a cluttered-looking or confusing visualization. More exploration on visualization methods is necessary for something this specific. Another obvious extension is to add movements that include both legs.

Finally, we are interested in integrating Limb-O into a virtual or augmented reality interface. Instead of attempting to interpret three-dimensional data through changes in camera angle on a two-dimensional screen, the user would be able to exist in the same space as these orbs for a more immersive experience. It would be interesting to also test the psychological effects on learning when being immersed in the motion visualization. Studies on the level of experienced immersion in VR with controllable virtual legs are largely non-existent outside of using virtual legs to help treat phantom limb pain [1].

6 CONCLUSION

In this work we presented Limb-O, a real-time solution for lower-limb motion comparison and feedback. A user validation study was run and we found that Limb-O outperformed standard video comparison both quantitatively and qualitatively. Users reported liking the score-based feedback provided by Limb-O, and some reported feeling less confused during the learning process due to real-time feedback confirming when the motion was being replicated correctly. One user also reported that "it was fun achieving better scores after practicing again and again".

Our study showed that while users found motion replication with Limb-O to be more uncomfortable and physically awkward, they also replicated the motions more accurately. This shows that, like a real coach, Limb-O was able to push users past their point of comfort in order to better master the reference motions. Our results also showed that users had more fun using Limb-O and were thus more likely to use it again in the future over a video comparison method for learning, boding well for the idea of implementing gamification in a virtual coach app.

As technology continues to become more accessible, we anticipate virtual coaching applications like Limb-O increasing in popularity. With further research in this area we hope that one day they will be effective enough to replace the need for in-person lessons. Further development of such accessible and affordable learning applications have the potential to help overcome the unfortunate financial and social barriers to education faced by many people.

7 ACKNOWLEDGEMENTS

We acknowledge financial support from an NSERC Strategic Project Grant to Menon, and CRC funding to Irani.

REFERENCES

[1] Elisabetta Ambron, Alexander Miller, Katherine J. Kuchenbecker, Laurel J. Buxbaum, and H. Branch Coslett. 2018-02-19. Immersive Low-Cost Virtual Reality Treatment for Phantom Limb Pain: Evidence from Two Cases.(Report). *Frontiers in Neurology* 9 (2018-02-19), 67.

[2] Guha Balakrishnan, Frédo Durand, and John Guttag. 2015. Video diff: highlighting differences between similar actions in videos. *ACM Transactions on Graphics* 34, 194 (2015).

[3] Peggy E. Browning. 2013. Hinge Joint. , 1634,1635 pages.

[4] Jim Buckley, Tabea DeWille, Chris Exton, Geraldine Exton, and Liam Murray. 2018. A Gamification–Motivation Design Framework for Educational Software Developers. *Journal of Educational Technology Systems* 47, 1 (2018), 101–127.

[5] Jacob Cohen. 1988. *Statistical power analysis for the behavioral sciences* (2nd ed. – ed.). L. Erlbaum Associates, Hillsdale, N.J.

[6] Yue-Yung Hu, Laura M Mazer, Steven J Yule, Alexander F Arriaga, Caprice C Greenberg, Stuart R Lipsitz, Atul A Gawande, and Douglas S Smink. 2017-04-01. Complementing Operating Room Teaching With Video-Based Coaching. *JAMA Surgery* 152, 4 (2017-04-01), 318,325.

[7] Tara E Lambert, Lisa A Harvey, Christos Avdalis, Lydia W Chen, Sayanthinie Jeyalingam, Carin A Pratt, Holly J Tatum, Jocelyn L Bowden, and Barbara R Lucas. 2017. An app with remote support achieves better adherence to home exercise programs than paper handouts in people with musculoskeletal conditions: a randomised trial. *Journal of Physiotherapy* 63, 3 (2017), 161 – 167. https://doi.org/10.1016/j.jphys.2017.05.015

[8] J Lieberman and C Breazeal. 2007. TIKL: Development of a Wearable Vibrotactile Feedback Suit for Improved Human Motor Learning. *IEEE Transactions on Robotics* 23, 5 (2007), 919,926.

[9] Zoe Marquardt, João Beira, Natalia Em, Isabel Paiva, and Sebastian Kox. 2012. Super Mirror: a kinect interface for ballet dancers. In *CHI EA '12, Extended Abstracts on Human Factors in Computing Systems.* ACM, 1619–1624.

[10] Jorge Piazentin Ono, Arvi Gjoka, Justin Salamon, Carlos Dietrich, and Claudio T. Silva. 2019. HistoryTracker: Minimizing Human Interactions in Baseball Game Annotation. In *CHI '19, Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems.* ACM.

[11] Yogendra Patil. 2017. A Multi-interface VR Platform For Rehabilitation Research. In *CHI EA '17, Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems.* ACM, 154–159.

[12] R.M.C. Segundo, M.N. De Amorim, and C.A.S. Santos. 2017. CrowdSync: User generated videos synchronization using crowdsourcing. In *2017 International Conference on Systems, Signals and Image Processing (IWSSIP).* ACM, 1–5.

[13] Keisuke Shiro, Kazme Egawa, Takashi Miyaki, and Jun Rekimoto. 2019. InterPoser: Visualizing Interpolated Movements for Bouldering Training. In *CHI EA '19, Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems.* ACM.

[14] Atima Tharatipyakul. 2017. Supporting Visual Temporal Media Comparison. In *CHI '17, extended abstracts on human factors in computing systems.* ACM, 330–334.

[15] Atima Tharatipyakul. 2018. Towards a better video comparison: comparison as a way of browsing the video contents. In *OzCHI '18, Proceedings of the 30th Australian Conference on Computer-Human Interaction.* ACM, 349–353.

[16] Emanuel Todorov, Reza Shadmehr, and Emilio Bizzi. 1997. Augmented Feedback Presented in a Virtual Environment Accelerates Learning of a Difficult Motor Task. *Journal of Motor Behavior* 29, 2 (1997), 147–158.

[17] Yaron Ukrainitz and Michal Irani. 2006. Aligning Sequences and Actions by Maximizing Space-Time Correlations. In *Computer Vision – ECCV 2006.* ACM, 538–550.

[18] Frederik Wiehr, Felix Kosmalla, Florian Daiber, and Antonio Krüger. 2016. betaCube: Enhancing Training for Climbing by a Self-Calibrating Camera-Projection Unit. In *CHI EA '16, Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems.* ACM, 1998–2004.