

CameraReady: Assessing the Influence of Display Types and Visualizations on Posture Guidance

Hesham Elsayed, Phillipp Hoffmann, Sebastian Günther, Martin Schmitz, Martin Weigel, Max Mühlhäuser, Florian Müller

2021

Preprint:

This is an accepted article published in Designing Interactive Systems Conference 2021 (DIS '21). The final authenticated version is available online at: <https://doi.org/10.1145/3461778.3462026>

CameraReady: Assessing the Influence of Display Types and Visualizations on Posture Guidance

Hesham Elsayed
TU Darmstadt
Darmstadt, Germany

Martin Schmitz
TU Darmstadt
Darmstadt, Germany

Philipp Hoffmann
TU Darmstadt
Darmstadt, Germany

Martin Weigel
Honda Research Institute Europe
Offenbach am Main, Germany

Sebastian Günther
TU Darmstadt
Darmstadt, Germany

Max Mühlhäuser
TU Darmstadt
Darmstadt, Germany

Florian Müller
TU Darmstadt
Darmstadt, Germany

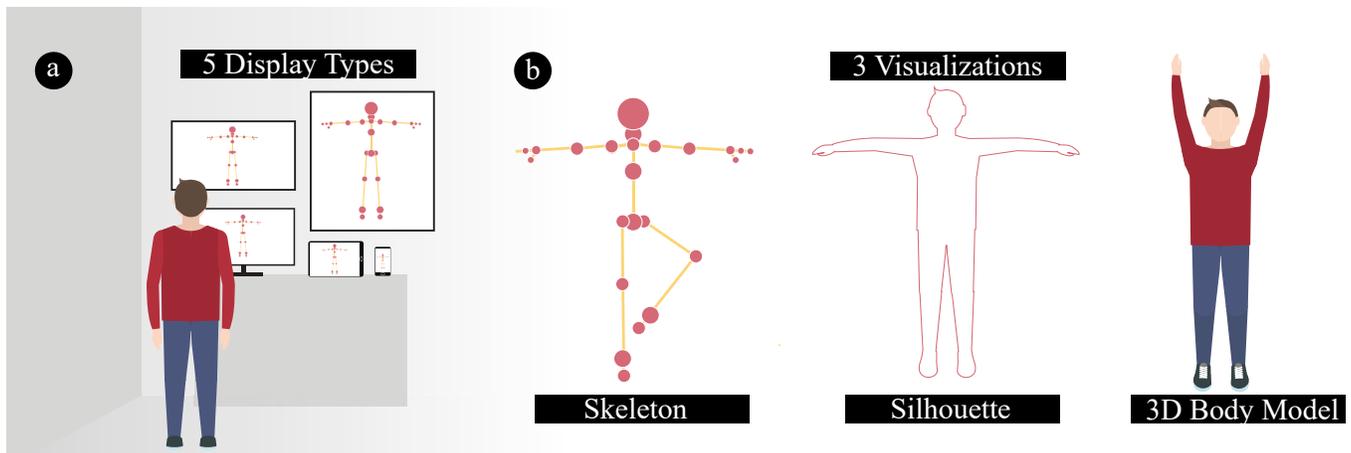


Figure 1: We introduce CameraReady, a mobile and cross-platform posture guidance system. We use CameraReady to evaluate the influence of (a) five different display types and (b) three visualizations on user performance.

ABSTRACT

Computer-supported posture guidance is used in sports, dance training, expression of art with movements, and learning gestures for interaction. At present, the influence of display types and visualizations have not been investigated in the literature. These factors are important as they directly impact perception and cognitive load, and hence influence the performance of participants. In this paper, we conducted a controlled experiment with 20 participants to compare the use of five display types with different screen sizes: smartphones, tablets, desktop monitors, TVs, and large displays. On each device, we compared three common visualizations for posture guidance: skeletons, silhouettes, and 3d body models. To conduct our assessment, we developed a mobile and cross-platform system

that only requires a single camera. Our results show that compared to a smartphone display, larger displays show a lower error (12%). Regarding the choice of visualization, participants rated 3D body models as significantly more usable in comparison to a skeleton visualization.

CCS CONCEPTS

• **Human-centered computing** → **Interactive systems and tools; Empirical studies in HCI.**

KEYWORDS

posture guidance, visualization, display, augmented reality

Permission to make digital or hard copies of part or all 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 third-party components of this work must be honored. For all other uses, contact the owner/author(s).
DIS '21, June 28-July 2, 2021, Virtual Event, USA
© 2021 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-8476-6/21/06.
<https://doi.org/10.1145/3461778.3462026>

ACM Reference Format:

Hesham Elsayed, Philipp Hoffmann, Sebastian Günther, Martin Schmitz, Martin Weigel, Max Mühlhäuser, and Florian Müller. 2021. CameraReady: Assessing the Influence of Display Types and Visualizations on Posture Guidance. In *Designing Interactive Systems Conference 2021 (DIS '21), June 28-July 2, 2021, Virtual Event, USA*. ACM, New York, NY, USA, 10 pages. <https://doi.org/10.1145/3461778.3462026>

1 INTRODUCTION

Traditionally, users learn new movements by following instructions of an experienced coach. While this approach has proven to be effective [31], it depends on the availability of the coach and the user at the same time and place, and is limited by the attention span of coach/user and high costs. Moreover, during the recent pandemic's time, traditional training with a coach becomes even more unlikely.

Recent advances in sensing and actuation technologies have led to the development of a wide-range of posture guidance applications [1, 3, 15, 17, 19, 21, 25, 35, 39, 42] to alleviate the limitations of the traditional approach. Sports training [16, 38], physiotherapy [35], martial arts [18], dance training [8, 29] and interacting with user interfaces [40] have been supported by digital posture guidance.

Smartphones and tablets carry sensors that can be leveraged, e.g., to adjust user balance [14]. While research has investigated mobile sensors for input [14, 37], it remains unclear if the visual output of such devices that is limited (e.g., small display size) is suitable for posture guidance. Desktop monitors and television sets have been shown to be useful, but commodity devices do not offer the sensing capabilities required. Immersive displays can be used to visualize movements in first and third person perspectives. However, transferring movements learned from VR to the real world is limited [26].

In this work, we assess the influence of display types and visualizations on user performance. To this end, we present a mobile system with a simple setup that works cross-platform. We used our system to conduct a controlled experiment in a lab environment with 20 participants, where we used five displays (smartphones, tablet, desktop monitor, TV, and large display) commonly found on the consumer market and three visualizations (skeletons, silhouettes, and 3D body models) commonly found in the literature. As part of our results, we identified that posture guidance is more accurate and more usable using larger displays and that 3D body models are more usable in comparison to a skeleton visualization.

In summary, the contributions of this paper are two-fold:

- A mobile, cross-platform system for posture guidance. This system, called CameraReady, enables scientific research on posture guidance across a wide variety of devices. We contribute this system as an open source framework to accelerate future research in that domain.
- A controlled experiment to evaluate the influence of display types and visualizations on user performance. The findings show that larger displays lead to higher user accuracy in comparison to a smartphone and have higher usability scores. Furthermore, the use of a 3D body model as visualization leads to higher usability scores compared to a skeleton visualization.

2 RELATED WORK

This work relates to motion guidance, display types, visualizations, and motion capture systems. In the following, we discuss approaches for motion guidance and their limitations, the various displays and visualizations currently used, as well as systems used for motion capture.

2.1 Motion Guidance

A large body of research exists on supporting users while learning new movements [1, 17, 19, 25, 35]. Use-cases range from physiotherapy [35], sports [16, 18, 38] and dance training [8, 29] to learning gestures for interaction [6, 32, 40]. In the traditional setting, users are supported by a coach. This enhances the learning experience by targeting these factors: (1) observational practice, (2) the user's focus of attention, (3) feedback and self-controlled practice [31]. However, due to the limitations of the traditional approach, such as requiring the presence of both coach and user, high costs for the users, limited attention span of coach (worsened in a group learning scenario), approaches have been proposed to overcome these problems. Recordings of training sessions help users learn new movements at home, but suffer from a lack of control over feedback and the user's focus of attention. Combining sensing technology for input and visual displays for output has the potential to overcome these issues by providing adequate feedback and guiding the user's attention appropriately. In this work, we assess the influence of display types and visualizations on posture guidance applications.

2.2 Display Types

This work investigates the effect of display type on posture guidance. Among the most common display types found on the consumer market are (1) smartphones, (2) tablets, (3) desktop monitors, (4) TVs and (5) large displays. In the following we discuss work done on posture guidance using these devices.

Smartphones. At present, little or no work is available that utilizes capabilities of modern smartphones for posture guidance. Research projects use the smartphones built-in accelerometer and gyroscope sensors, in combination with audio instructions to help users adjust their posture [14, 37]. This helps, for instance, for rehabilitation by improving balance of users and hence their stance. Products on the market use smartphones synchronously with wearable sensors for ergonomics, e.g. UPRIGHT GO¹. Very little research has been done on using visual input and output capabilities currently present in modern smartphones. Start-ups, such as Onyx², use the smartphone's camera to aid users in their training by providing statistics on simple exercises, but offer no digital guidance. In this research we use the RGB camera of the smartphone for tracking the user's posture. More advanced camera options like depth-cameras are being introduced in some smartphones, but are not yet widely available.

Tablets. Similar to smartphones, tablets carry accelerometers and gyroscope sensors that can be used to help users with exercises, e.g. Plankpad³. However, tablets cannot be stored in pockets. This limits their use in applications, such as balance improvement [14] or improvement of golf swings [37]. On the other hand, tablets offer a larger screen size that can be used to help users during full-body posture guidance.

Desktop Monitors. Platforms such as YouTube⁴ have made workout routines accessible on all types of displays, e.g. POPSUGAR

¹<https://www.uprightpose.com>

²<https://www.onyx.fit/>

³<https://plankpad.com/>

⁴<https://www.youtube.com/>

	Anderson et al. [1]	Badler et al. [3]	Bailenson et al. [4]	Barioni et al. [5]	Chan et al. [8]	Chua et al. [9]	Da Gama et al. [10]	Dürr et al. [11]	Elsayed et al. [12]	Fourie and van der Haar [13]	Hamanishi et al. [16]	Han et al. [17]	Han et al. [18]	Hoang et al. [19]	Hülsmann et al. [22]	Kallmann et al. [24]	Katzakis et al. [25]	Lin et al. [27]	Marquardt et al. [29]	Sodhi et al. [32]	Sousa et al. [33]	Tadayon et al. [34]	Tang et al. [35]	Tao et al. [36]	Veloso et al. [38]	Wei et al. [41]	Yang and Kim [42]	Walter et al. [40]	
Visualization	⚙️	⚙️	⚙️	⚙️	⚙️	⚙️	⚙️	⚙️	⚙️	⚙️	⚙️	⚙️	⚙️	⚙️	⚙️	⚙️	⚙️	⚙️	⚙️	⚙️	⚙️	⚙️	⚙️	⚙️	⚙️	⚙️	⚙️	⚙️	⚙️
Display	📺	📺	📺	📺	📺	📺	📺	📺	📺	📺	📺	📺	📺	📺	📺	📺	📺	📺	📺	📺	📺	📺	📺	📺	📺	📺	📺	📺	📺
Mocap	📷	📷	📷	📷	📷	📷	📷	📷	📷	📷	📷	📷	📷	📷	📷	📷	📷	📷	📷	📷	📷	📷	📷	📷	📷	📷	📷	📷	📷

Table 1: Overview of posture guidance approaches using skeletons ⚙️, silhouettes 📷 and avatars 📺 for visualizations; life-sized displays 📺, such as public displays and immersive displays, and desktop monitors 📺 for output; and depth sensors 📷, wearable sensors 📷, camera-based 📷 and marker-based 📷 motion capture for input.

Fitness ⁵. They are often watched on desktop monitors, due to their larger size compared to mobile devices. While using this approach has little or no hurdles for many users (a desktop monitor and an internet connection), video content lacks on user feedback, which is crucial while learning to perform new movements correctly. Most of today’s desktop monitors can be easily extended with a webcam for input, which enables the posture tracking applications envisioned in this paper.

Television Sets. TV sets are suitable for full-body workouts at home due to their large screen sizes. Similar to desktop monitors, approaches that use a TV for posture guidance, e.g using video feed of a trainer, lack on user feedback. Research has, therefore, introduced approaches that overcome this limitation, e.g. for physiotherapy [35].

Wall-sized & Immersive Displays. As a consequence of the high potential impact of posture guidance on many of applications, a lot of research has been done looking into interaction techniques with public displays, virtual and augmented environments. Public displays help users learn gestures for interaction [40]. VR and AR enable posture guidance using both first and third person perspectives [17, 19]. Currently, transferring movements learned in VR to the real world [26], as well as the use of different visualizations is not well understood.

Current approaches for posture guidance are designed for use in a particular setup with one type of display. Although most approaches target full-body posture guidance, some approaches, e.g. [17], aim at guiding only limb movements. With CameraReady, we introduce a cross-platform system for multiple display sizes that offers full-body posture guidance, is self-contained, mobile and person independent.

2.3 Visualizations

In the following, visualizations commonly found in the literature are discussed. An overview of the visualizations and related work can be seen in Figure 2 and Table 1, respectively.

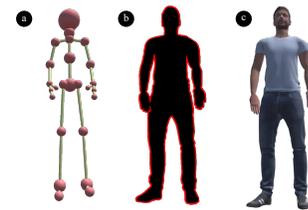


Figure 2: Visualizations used for posture guidance: (a) skeleton, (b) silhouette and (c) 3D body model.

Skeletons. A skeleton visualizations offers a simple, abstract representation of the joints in the human body. Joints are connected by abstract representations of bones. By using this body representation, details of body composition and deformation are lost. These can serve as additional indicators for depth information, which is important when using a 2D screen to display a 3D body posture. On the other hand, by abstracting away information a user could perceive a body posture faster. An overview of research using skeletons as a visualization for posture guidance can be found in Table 1.

Silhouettes. A silhouette visualization shows the outline of the human body. Information about joint locations, body composition and deformation are not displayed. This might make perceiving depth information even more challenging in comparison to a skeleton visualization. Examples of research using silhouettes can be seen in Table 1.

3D Body Models. A 3D body model offers the most detailed representation of the human body among the mentioned categories. This visualization can be very realistic, e.g. by matching body composition [28] and deformation [30] during motion, or less realistic, e.g., by using a virtual avatar Barioni et al. [5]. It is unclear how the differences between these visualizations influence posture guidance applications.

⁵<https://www.youtube.com/channel/UCBINFWq52ShSgUFEoyfnSwg>

2.4 Motion Capture

Approaches for motion sensing can be categorized into (1) marker-based, (2) markerless, and (3) wearable sensors. Next, advantages and disadvantages of these categories are discussed.

Marker-based. Marker-based approaches rely on the use of retro-reflective markers and a set of high speed cameras for precise motion capture. Marker-based systems are affected by occlusion, not mobile and require the user to wear markers. Optitrack⁶ and Vicon⁷ offer marker-based motion capture systems.

Markerless. Classically, human posture estimation using camera images has been accomplished by feature engineering and extraction from images [15]. More recently, approaches using deep neural networks have been introduced to overcome some of the problems of hand-crafted features used by, e.g. depth cameras. Camera-based methods, however, are affected by occlusion, are computationally demanding and in cases where a more complex setup, e.g., The Captury⁸ is required, mobility is reduced.

Wearable Sensors. Using wearable sensors, human posture estimation free from occlusion problems can be achieved [3, 21, 39, 42]. These systems are also usually highly mobile, but suffer from problems such as drift when using inertial measurement units (IMU) and require users to wear bulky equipment. An example provider of motion capture systems with wearable sensors is Xsens⁹.

2.5 Summary

A multitude of display types are available to support the development of posture guidance applications. Hurdles that stop the use of ubiquitous smartphones include the small screen size and limited computation power. However, little research has investigated the use of these devices for this purpose. Desktop monitors and TV sets have been used previously, but are limited in their sensing capabilities. Immersive displays, in particular, VR environments are limited in their capacity to teach users new movements that are transferable to the real world [26]. Common to all displays, the influence of the visualization used is not well understood.

3 CAMERAREADY

In the following, we describe our mobile cross-platform system for posture guidance on various display types. The system is motivated by three example application scenarios, which we use to elicit five design requirements. Afterwards, we discuss our implementation, the neural network we use for pose estimation, and evaluate the accuracy of this approach.

3.1 Example Application Scenarios

We present a set of example application scenarios that show the utility of CameraReady for motion guidance.

3.1.1 Remote Motion Coaching. One of the major disadvantages of traditional coaching is the necessity of physical presence. With CameraReady a coach can remotely teach users in real-time with

the possibility of offering feedback (Figure 3a). In comparison to a pre-recorded video feed of the coach, this approach allows active support during the learning process. This also makes it possible to make use of algorithms for automatic classification of errors [23].

3.1.2 Computer-supported Mobile Training. In addition to real-time support by a coach, CameraReady can use automated algorithms to support users in practicing new and previously learned movements in the comfort of their homes and outside (Figure 3b). This can be used for example while practicing Yoga for showing correct postures and highlighting errors in the user's posture. In comparison to remote motion coaching, the focus here is not on supporting a 2-way communication channel between a coach and a user, but on supporting mobile training scenarios without necessarily requiring a coach.

3.1.3 Ergonomics. While interacting with displays, ergonomics is a very important factor [2]. CameraReady can augment the display with information on incorrect postures and guide users in performing necessary adjustments to prevent short- and long-term injuries (Figure 3c).

3.2 Design Requirements

Based on our example application scenarios we extracted the following five design requirements for our system:

3.2.1 Full-Body Guidance. Most movements in sports, martial arts, and dance require movements of the full body. Hence, our system needs to give guidance for full-body postures. In addition, some modes could be supported for specialized training of a specific body part by ignoring movement of the other body parts.

3.2.2 Cross-platform. Many people own or have access to a wide variety of digital devices, including smartphones, tablets, notebooks, and TVs, which can be used for computer-supported mobile training. The system should work on a large variety of these devices to allow the user to chose his/her preferred training setup. For sensing, the system should require no special equipment, such as optical tracking systems or depth cameras. Instead, the system should work with a single-camera setup, which is already included in most commercial devices.

3.2.3 Self-Contained. The system should consist of a single device and not require a complex setup, especially in mobile training scenarios. Moreover, it should avoid attachment of wearable sensors on the user, since the additional weight might change the body perception and could make some movements more difficult to execute and learn.

3.2.4 Mobile. The posture guidance system should be mobile to allow for assistance in many environments, e.g. outside in a park or in a fitness center. Hence, the system should work on smaller-sized displays and should be able to work battery operated.

3.2.5 Person Independent. In some of our envisioned setups the device might be shared by multiple people. For example, a television might be used by all members of a family and an ergonomics trainer for a machine could support factory workers. Hence, the system should not be personalized to a specific user. This also allows multiple users to share the same system. For example, a in-person

⁶<https://optitrack.com/>

⁷<https://www.vicon.com/>

⁸<https://thecaptury.com>

⁹<https://www.xsens.com/>

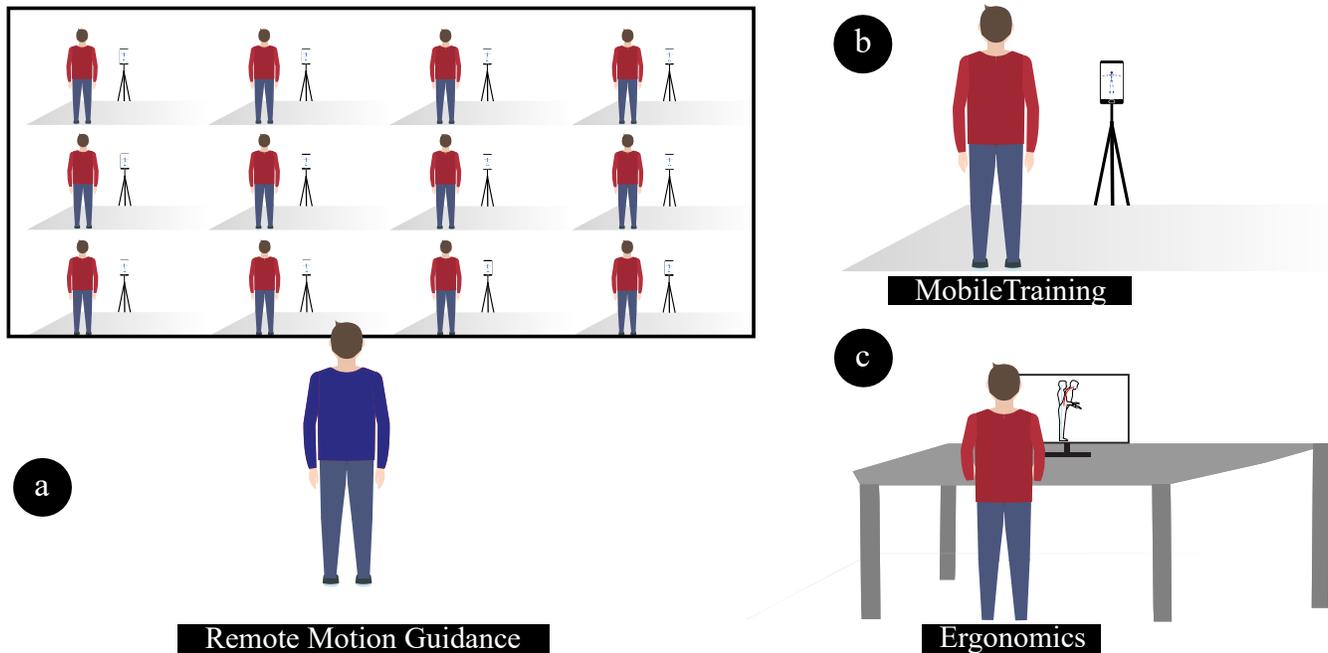


Figure 3: Example application scenarios of CameraReady.

training could use the system to show the specifics of a posture, before asking the student to imitate the movement.

3.3 Implementation

Our implementation of CameraReady consists of two parts. First, a client application running on Android or Windows 10, and second, a server application running on Windows 10. Client applications were responsible for sending the camera feed to the server for evaluation and displaying the visualizations. We used Unity distribution 2019.3.15f1 for compatibility between client-server applications. Unity Neural Networking Package Barracuda¹⁰ was used for running the neural network, uTextureSendReceive¹¹ for handling camera feed and Unity Mirror Networking¹² for synchronisation of variables, game object positions and network messages. We open-source CameraReady to accelerate future research in the area of visual posture guidance on Github¹³.

3.4 Neural Network Evaluation

We used a neural network implemented by Yukihiro Aoyagi¹⁴ for real-time (up to 60 fps) pose estimation. This approach uses MobileNetV3 [20]. We chose this network over OpenPose [7] as it detects human postures in 3D without requiring a setup with multiple cameras. We further conducted an evaluation of the performance of this network on 3D joint angle error. To do so, the postures used during the experiment were performed while wearing retro-reflective markers and being tracked by an Optitrack V100:R2

motion capture system. Additionally, the devices used during the experiment captured images of the postures. A total of 60 postures and images were collected: 20 per smartphone, tablet, and webcam. Estimation of the neural network was compared to ground truth data collected from Optitrack. Figure 4a shows the setup and Figure 4b-d show the results of our evaluation. The neural network had an average error of 5.57° over all joints, postures and devices used making it comparable to a Microsoft Kinect [36].

4 EXPERIMENT

We conducted a controlled lab experiment to investigate the influence of different visualizations and display types on the accuracy, efficiency and usability of posture guidance. In particular, we aim to answer the following hypotheses:

- H1** Larger displays lead to higher user accuracy.
- H2** Larger displays lead to lower task completion time (TCT).
- H3** Larger displays lead to higher usability ratings.
- H4** Visualizations with a higher level of detail lead to higher user accuracy.
- H5** Visualizations with a higher level of detail lead to lower TCT.
- H6** Visualizations with a higher level of detail lead to higher usability ratings.

4.1 Participants

We recruited 20 participants (13 Male, 5 Female, 1 Diverse and 1 participant did not specify) aged between 21 and 43 years ($M = 25.25$, $SD = 5.30$). 15 participants are practicing or had prior experience with posture training exercises. Participation in our experiment was voluntary and no compensation was offered. There were some sweets in the lab, if participants wanted to have some.

¹⁰<https://docs.unity3d.com/Packages/com.unity.barracuda@0.3/manual/index.html>

¹¹<https://github.com/BarakChamo/uTextureSendReceive>

¹²<https://mirror-networking.com/>

¹³<https://gitlab.com/ph-industries/CameraReady>

¹⁴<https://github.com/digital-standard/ThreeDPoseTracker>

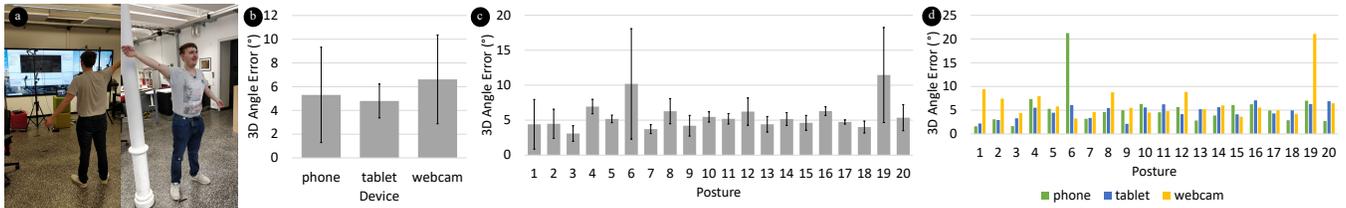


Figure 4: (a) Setup used in our neural network evaluation, (b) 3D angle errors across devices, (c) 3D angle errors per posture, and (d) 3D angle errors for devices and postures combined.

4.2 Design

We used a within-subject design with *display type* (smartphone, tablet, desktop monitor, TV, and large display) and *visualization type* (skeletons, silhouettes and 3D body models) as independent variables. We counterbalanced the order of *display type* and *visualization type* using a 5x10 and a 3x6 balanced Latin square. Displays were shown in the order determined by the 5x10 balanced Latin square. Within each display condition, the order of visualizations was taken from the 3x6 Latin balanced square. Remaining visualization combinations were transferred to the next display condition. By using this approach, we conducted the experiment with 20 participants instead of 30. For each combination of levels of independent variables, participants performed 20 postures from a baseline dataset [12] in randomized order. Figure 5 shows the postures. This resulted in a total of 300 postures per participant.

4.3 Procedure

To prevent the spread of COVID-19, all materials used in the experiment were disinfected and the lab was aired for a minimum of 30 minutes between participants. After welcoming and obtaining informed consent from our participants, we collected their demographic data. Then, we explained the task and provided a brief overview of the procedure. The task was to stand at a particular location looking at a display and imitate the visible posture. After each condition, participants filled out a System Usability Scale (SUS) questionnaire. Finally, participants filled out a questionnaire asking for their favorite display type and visualization and their reasoning. The questionnaire also asked for feedback and comments on the overall experience and suggestions for improving the system.

4.4 Apparatus

We used a Zotac Magnus EN1060K all-in-one computer with an NVIDIA GeForce GTX 1060 graphics card while running our experiment. During the experiment participants had to signal that they are ready to start, and confirm that they performed the required posture. Interactions of the participants with the system were performed by pressing any button on a Logitech Presenter R400. We used five commonly available devices to represent the five display types:

- *Smartphone*: Google Pixel 3 with a 5.5" screen (2,280 x 1,080). We used the ultra wide front camera with a focal length of 2.03mm (2448 x 3264). The phone was placed 1.20m away from the participant. This distance was determined from

pilot tests as the distance where a 1.80m tall person is fully visible with limbs stretched.

- *Tablet*: Microsoft Surface Pro 4 with a 12.3" screen (2,736 x 1,824). We used the built-in Intel AVStream 2500 camera (2560 x 1440) with a x0.62 wide lens to increase field of view. Participants stood 2.0m away from the tablet, in order to be fully visible.
- *Desktop monitor*: HP 24" display (1920 x 1080) and a Logitech QuickCam Pro 9000 (1600 x 1200) with a 3.7mm focal length. 3.5m distance was required between participants and the webcam.
- *TV*: We used a NEC Multisync UN551VS 55" display (1920 x 1080) for the TV condition, and a Logitech QuickCam Pro 9000 was used for input.
- *Large display*: We used two NEC Multisync UN551VS 72" vertically stacked for the large display condition and a Logitech Quickcam Pro 9000 was used for input.

4.5 Data Analysis

We analyzed the recorded data using a two-way repeated measures ANOVA with *display type* and *visualization type* as the two independent factors. We tested the data for normality with Shapiro Wilk's test and found no significant deviations. Where Mauchly's test indicates a violation of the assumption of sphericity, we corrected the tests using the Greenhouse-Geisser method and report the ϵ . When significant effects are revealed, we use Bonferroni corrected pairwise t-tests for post-hoc analysis.

5 RESULTS

This section details on the results of varying our independent variables *display type* and *visualization type* on the dependent variables error, Task Completion Time (TCT), and SUS scores.

5.1 Error

We analysed the average 3D joint angle error of participants after imitating postures. A 2-way repeated measures ANOVA showed a significant main effect of *visualization* ($F_{1.75, 31.52} = 3.94$, $p < .05$, $\epsilon = 0.88$, $\eta^2 = 0.005$) and *display* ($F_{1.58, 28.44} = 5.26$, $p < .05$, $\epsilon = 0.40$, $\eta^2 = 0.10$). Post-hoc tests revealed no significant differences between skeletons ($M = 22.18^\circ$, $SE = 4.26^\circ$), silhouettes ($M = 22.25^\circ$, $SE = 4.47^\circ$) and 3D body models ($M = 21.73^\circ$, $SE = 4.22^\circ$); while confirming significant differences between smartphone ($M = 24.06^\circ$, $SE = 3.91^\circ$) and desktop monitors ($M = 21.05^\circ$, $SE = 4.54^\circ$, $p < .01$), TVs ($M = 21.22^\circ$, $SE = 4.49^\circ$, $p < .05$) and large displays ($M = 21.07^\circ$, $SE = 3.88^\circ$,

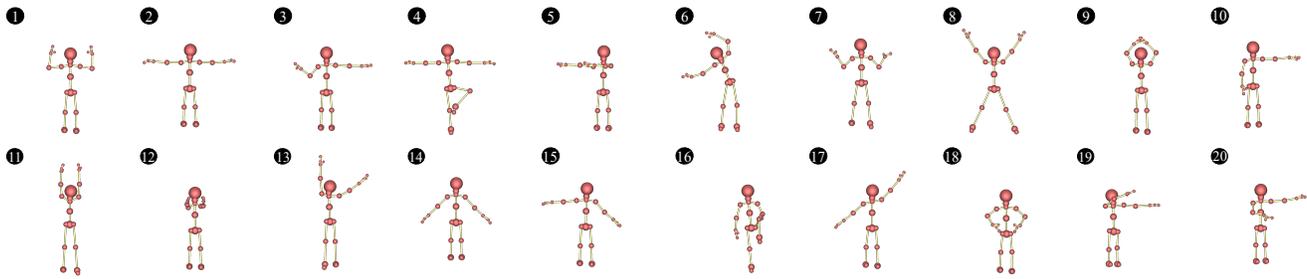


Figure 5: Postures used in our experiment.

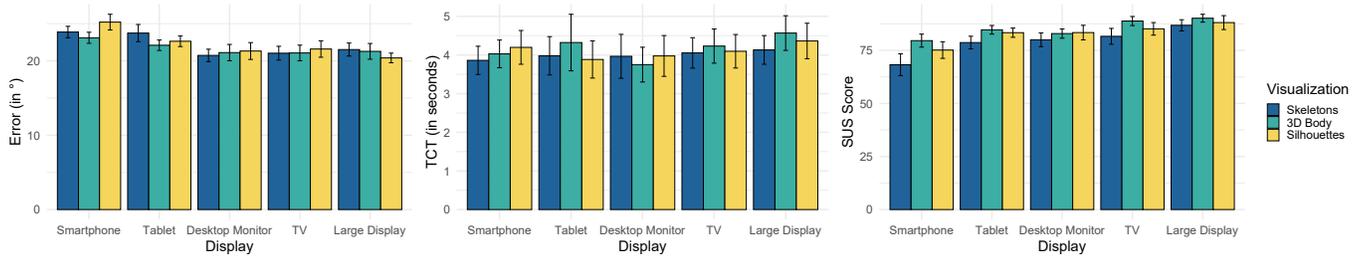


Figure 6: Error, TCT, and SUS scores of users across visualizations and displays used. Error bars are the standard errors.

$p < .01$). We found no interaction effects between *visualization* and *display* ($F_{3,00,53.92} = 2.62$, $p > .05$).

5.2 Task Completion Time (TCT)

We measured TCT as the time participants took to transition from a neutral to the displayed posture. A 2-way repeated measures ANOVA showed no significant main effects for both *visualization* ($F_{2,36} = 1.42$, $p > 0.05$) and *display* ($F_{4,72} = 0.60$, $p > 0.05$). Participants were on average quicker with skeletons ($M = 4.00s$, $SE = 1.96s$), than silhouettes ($M = 4.10s$, $SE = 2.05s$) and 3D body models ($M = 4.18s$, $SE = 2.22s$). The desktop monitor display had the lowest TCT on average ($M = 3.90s$, $SE = 2.28s$), followed by smartphone ($M = 4.03s$, $SE = 1.71s$), tablet ($M = 4.06s$, $SE = 2.56s$), TV ($M = 4.13s$, $SE = 1.85s$), and large displays ($M = 4.35s$, $SE = 1.87s$).

5.3 System Usability Scale (SUS)

After each condition, we assessed usability of our system by asking participants to fill out a SUS. A SUS score ranges from 0 - 100, with higher scores showing higher usability. We found significant main effects for *visualization* ($F_{1,48,16.33} = 5.37$, $p < .05$, $\epsilon = 0.74$, $\eta^2 = 0.06$) and *display* ($F_{1,55,17.10} = 6.41$, $p < .05$, $\epsilon = 0.39$, $\eta^2 = 0.164$) on SUS scores. Post-hoc tests showed a significant difference between skeletons ($M = 79.16$, $SE = 15.91$) and 3D body models ($M = 85.15$, $SE = 10.19$, $p < .05$). Usability of smartphone ($M = 74.43$, $SE = 17.44$) compared to tablet ($M = 82.21$, $SE = 10.21$, $p < .05$), TV ($M = 85.19$, $SE = 12.66$, $p < 0.01$), and large display ($M = 88.32$, $SE = 10.94$, $p < 0.001$) was also significant. No interaction effects between *visualization* and *display* ($F_{3,25,35.80} = 2.57$, $p > .05$) were found.

5.4 User Feedback

In the final questionnaire, we asked participants which visualizations and displays they found most appealing and collected comments on additional features from our users.

5.4.1 Preferences & Perceptions. A 3D body model was favoured by 40% of our participants as they found posture to be "easy to recognize" (P12), "clearer in comparison to other visualizations" (P5), and "most visible" (P10). 30% of the participants preferred skeletons as they "covered the least of my body and allowed me to estimate how I was positioned" (P1), had a "good mixture between transparency and overlay" (P3) and other visualizations had "a different shoulder position" (P2). Remaining participants preferred silhouettes as they were "most visible" (P16) and offered the "easiest (way) to see what to do" (P11). Large displays were the favorite display type by 90% of our users as they "made the recognition of the displayed poses effortless" (P1), were "similar to an expert showing you an exercise" (P3), and offered an improvement in comparison to other displays "the smaller the screen size the harder to read the pose" (P15). The remaining 10% of our users preferred the use of TV as it was "very pleasant to use" (P13).

5.4.2 Suggested Features. In the end, we asked participants if they would like to see additional features in a posture guidance. Participants suggested a "color-coded avatar to show how well which joint imitates the shown pose" (P1) and using colors to clarify in the skeleton visualization "if joints are in front or behind other joints" (P2). One participant suggested a "transparency function to see your own pose better" (P3). The participants also expressed a need for "a display of the progress" (P13). The progress could be used as a "verification that the pose is well done" (P6) and "to see how many positions are still remaining [to increase motivation]" (P5). Participants further

suggested combining progress indication with *"visuo-auditory feedback to show that you have taken the right posture"* (P5). Participants proposed multi-view visualisations since *"poses that use depth are difficult to imitate"* (P14). For example, *"a second display that shows the pose from the side"* (P14) might be useful. Moreover, participants suggested *"animations how to get to a certain pose"* (P20). Lastly, participants suggested the use of a female avatar *"for a better alignment of posture"* (P5).

6 DISCUSSION

In this section, we discuss quantitative and qualitative results of our experiment. In general, our results show the feasibility of posture guidance using all displays and visualizations presented in this paper. We observed significant differences in usability scores and user accuracy across devices and visualizations. In addition to our quantitative results, users expressed clear preferences to larger devices.

6.1 Display Types

We found that larger devices were rated as more usable by participants. Therefore we can accept **H3**. Regarding accuracy of imitating postures, larger devices such as desktop monitors, TVs, and large displays showed lower errors in comparison to a smartphone (12.51%, 11.80% and 12.43%, respectively). These results support **H1** when comparing a smartphone to other devices. However, when comparing tablet, desktop monitor, TV and large display we cannot support **H1**. Contrary to our hypothesis, smartphones can be used for full body posture guidance. Although lower than large displays, smartphones received a SUS score of 74 and had comparable TCT. TCT was comparable across all displays, hence we cannot support **H2**.

6.2 Visualizations

Considering the influence of visualization on posture guidance, we found no significant differences that show superiority of a particular visualization. Our initial hypothesis (**H4**) was that more detailed visualizations lead to lower error, which was not reflected in our quantitative results but was mentioned by our participants. Depth information (i.e., how body parts are relatively arranged) was not clear to participants when using skeletons and silhouettes. Moreover, SUS scores showed 3D body models to be more usable than skeletons, hence we can accept **H6**. Regarding TCT, all visualizations had comparable performance. Therefore, we cannot support **H5**.

6.3 Subjective Preferences

In line with related work, users expressed the need for multi-view visualizations [35], color coding of misaligned body parts [19], and audio instructions [1]. Participants further commented on useful features, e.g., addition of transparency of the visualization to make a person's own posture more visible, color coding depth information, and animating the visualization to see the motion required.

7 LIMITATIONS

We are confident that our result provide valuable insights into the influence of different visualizations and display types on the accuracy and efficiency of posture guidance systems. However, design as well as the results of our experiment impose some limitations and starting points for future work.

7.1 Real-World Applicability

In this paper, we investigated the accuracy, efficiency and usability of posture guidance systems in a lab setting. We chose this approach to focus on the mere influence of the factors and to exclude external influences. While we are convinced that our results make a strong contribution to the future of such systems, we also acknowledge that other settings might yield other results. Therefore, further work is necessary to understand how these results are transferable to in-the-wild settings.

7.2 Real-Time Feedback

CameraReady uses various visualizations for posture guidance. This guidance is currently limited to static postures that communicate the target posture to the user. Future work should, therefore, investigate the addition of live feedback to further support the user, as well as animations that make the transition between postures more intuitive.

7.3 Recognition Accuracy

Lastly, we used a neural network for mocap. This approach is not as precise as a marker-based system, however opens the door to a variety of interesting application scenarios in HCI. We are confident that this limitation will be solved by advances in the area of computer vision.

8 CONCLUSION

We presented CameraReady, a mobile system with a simple setup that works cross-platform. We have assessed the feasibility of our system in a controlled lab experiment, using different visualizations found in the literature and various displays found on the consumer market. Our results indicate that CameraReady can be used for posture guidance across different screen sizes and visualizations. While larger displays and a 3D body model visualization show highest usability scores, other devices and visualizations also proved to be usable in terms of our participants' SUS scores and accuracy.

ACKNOWLEDGMENTS

This work was supported by the German Federal Ministry of Education and Research (BMBF) SWC "ITMAR" (01IS17050). We would further like to thank Simone Schroeder for helping with the figures.

REFERENCES

- [1] Fraser Anderson, Tovi Grossman, Justin Matejka, and George Fitzmaurice. 2013. YouMove: Enhancing Movement Training with an Augmented Reality Mirror. In *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology (UIST '13)*. Association for Computing Machinery, New York, NY, USA, 311–320. <https://doi.org/10.1145/2501988.2502045>
- [2] Myroslav Bachynskiy, Gregorio Palmas, Antti Oulasvirta, Jürgen Steimle, and Tino Weinkauff. 2015. Performance and Ergonomics of Touch Surfaces: A Comparative Study Using Biomechanical Simulation. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of

- Korea) (*CHI '15*). Association for Computing Machinery, New York, NY, USA, 1817–1826. <https://doi.org/10.1145/2702123.2702607>
- [3] Norman Badler, Michael Hollick, and John Granieri. 1993. Real-Time Control of a Virtual Human Using Minimal Sensors. *Center for Human Modeling and Simulation 2* (1993). <https://doi.org/10.1162/pres.1993.2.1.82>
 - [4] Jeremy Bailenson, Kayur Patel, Alexia Nielsen, Ruzena Bajcsy, Sang-Hack Jung, and Gregorij Kurillo. 2008. The Effect of Interactivity on Learning Physical Actions in Virtual Reality. *Media Psychology* 11, 3 (2008), 354–376. <https://doi.org/10.1080/15213260802285214>
 - [5] R R Barioni, W Costa, A Aleluia, and V Teichrieb. 2019. BalletVR: a Virtual Reality System for Ballet Arm Positions Training. In *2019 21st Symposium on Virtual and Augmented Reality (SVR)*. 10–16. <https://doi.org/10.1109/SVR.2019.00018>
 - [6] Olivier Bau and Wendy E Mackay. 2008. OctoPocus: A Dynamic Guide for Learning Gesture-based Command Sets. In *Proceedings of the 21st Annual ACM Symposium on User Interface Software and Technology (UIST '08)*. ACM, New York, NY, USA, 37–46. <https://doi.org/10.1145/1449715.1449724>
 - [7] Z. Cao, G. Hidalgo Martinez, T. Simon, S. Wei, and Y. A. Sheikh. 2019. OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2019).
 - [8] Jacky C P Chan, Howard Leung, Jeff K T Tang, and Taku Komura. 2011. A Virtual Reality Dance Training System Using Motion Capture Technology. *IEEE Trans. Learn. Technol.* 4, 2 (apr 2011), 187–195. <https://doi.org/10.1109/TLT.2010.27>
 - [9] Philo Tan Chua, Rebecca Crivella, Bo Daly, Ning Hu, Russ Schaaf, David Ventura, Todd Camill, Jessica Hodgins, and Randy Pausch. 2003. Training for Physical Tasks in Virtual Environments: Tai Chi. In *Proceedings of the IEEE Virtual Reality 2003 (VR '03)*. IEEE Computer Society, Washington, DC, USA, 87–. <http://dl.acm.org/citation.cfm?id=832289.835983>
 - [10] Alana Elza Fontes Da Gama, Thiago Menezes Chaves, Lucas Silva Figueiredo, Adriana Baltar, Ma Meng, Nassir Navab, Veronica Teichrieb, and Pascal Fallavolita. 2016. MirrARbilitation. *Comput. Methods Prog. Biomed.* 135, C (2016), 105–114. <https://doi.org/10.1016/j.cmpb.2016.07.014>
 - [11] Maximilian Dürr, Rebecca Weber, Ulrike Pfeil, and Harald Reiterer. 2020. EGuide: Investigating Different Visual Appearances and Guidance Techniques for Egocentric Guidance Visualizations. In *Proceedings of the Fourteenth International Conference on Tangible, Embedded, and Embodied Interaction (TEI '20)*. Association for Computing Machinery, New York, NY, USA, 311–322. <https://doi.org/10.1145/3374920.3374945>
 - [12] Hesham Elsayed, Martin Weigel, Julius von Willich, Markus Funk, and Max Mühlhäuser. 2019. APS: A 3D Human Body Posture Set as a Baseline for Posture Guidance. In *Proceedings of the 12th ACM International Conference on Pervasive Technologies Related to Assistive Environments (PETRA '19)*. Association for Computing Machinery, New York, NY, USA, 353–357. <https://doi.org/10.1145/3316782.3324012>
 - [13] Margaux Fourie and Dustin van der Haar. 2019. Ballet Pose Recognition: A Bag-of-Words Support Vector Machine Model for the Dance Training Environment. In *Information Science and Applications 2018*, Kuinam J Kim and Nakhoon Baek (Eds.). Springer Singapore, Singapore, 317–325.
 - [14] C. Franco, A. Fleury, P. Y. Gumery, B. Diot, J. Demongeot, and N. Vuillerme. 2013. iBalance-ABF: A Smartphone-Based Audio-Biofeedback Balance System. *IEEE Transactions on Biomedical Engineering* 60, 1 (2013), 211–215.
 - [15] Wenjuan Gong, Xuenan Zhang, Jordi González, Andrews Sobral, Thierry Bouwmans, Changhe Tu, and El-hadi Zahzah. 2016. Human Pose Estimation from Monocular Images: A Comprehensive Survey. *Sensors* 16, 12 (2016). <https://doi.org/10.3390/s16121966>
 - [16] Natsuki Hamanishi, Takashi Miyaki, and Jun Rekimoto. 2019. Assisting Viewpoint to Understand Own Posture as an Avatar In-Situation. In *Proceedings of the 5th International ACM In-Cooperation HCI and UX Conference (CHI-uXiD'19)*. Association for Computing Machinery, New York, NY, USA, 1–8. <https://doi.org/10.1145/3328243.3328244>
 - [17] Ping-Hsuan Han, Kuan-Wen Chen, Chen-Hsin Hsieh, Yu-Jie Huang, and Yi-Ping Hung. 2016. AR-Arm: Augmented Visualization for Guiding Arm Movement in the First-Person Perspective. In *Proceedings of the 7th Augmented Human International Conference 2016 (AH '16)*. ACM, New York, NY, USA, 31:1–31:4. <https://doi.org/10.1145/2875194.2875237>
 - [18] Ping-Hsuan Han, Yang-Sheng Chen, Yilun Zhong, Han-Lei Wang, and Yi-Ping Hung. 2017. My Tai-Chi Coaches: An Augmented-Learning Tool for Practicing Tai-Chi Chuan. In *Proceedings of the 8th Augmented Human International Conference on - AH '17*. <https://doi.org/10.1145/3041164.3041194>
 - [19] Thuong N Hoang, Martin Reinoso, Frank Vetere, and Egemen Tanin. 2016. One-body: Remote Posture Guidance System Using First Person View in Virtual Environment. In *Proceedings of the 9th Nordic Conference on Human-Computer Interaction (NordCHI '16)*. ACM, New York, NY, USA, 25:1–25:10. <https://doi.org/10.1145/2971485.2971521>
 - [20] Andrew Howard, Ruoming Pang, Hartwig Adam, Quoc Le, Mark Sandler, Bo Chen, Weijun Wang, Liang-Chieh Chen, Mingxing Tan, Grace Chu, Vijay Vasudevan, and Yukun Zhu. 2019. Searching for MobileNetV3. 1314–1324. <https://doi.org/10.1109/ICCV.2019.00140>
 - [21] Yinghao Huang, Manuel Kaufmann, Emre Aksan, Michael J Black, Otmar Hilliges, and Gerard Pons-Moll. 2018. Deep Inertial Poser: Learning to Reconstruct Human Pose from Sparse Inertial Measurements in Real Time. In *SIGGRAPH Asia 2018 Technical Papers (SIGGRAPH Asia '18)*. ACM, New York, NY, USA, 185:1–185:15. <https://doi.org/10.1145/3272127.3275108>
 - [22] Felix Hülsmann, Cornelia Frank, Irene Senna, Marc O Ernst, Thomas Schack, and Mario Botsch. 2019. Superimposed Skilled Performance in a Virtual Mirror Improves Motor Performance and Cognitive Representation of a Full Body Motor Action. *Frontiers in Robotics and AI* 6 (2019), 43. <https://doi.org/10.3389/frobt.2019.00043>
 - [23] Felix Hülsmann, Jan Göpfert, Barbara Hammer, Stefan Kopp, and Mario Botsch. 2018. Classification of motor errors to provide real-time feedback for sports coaching in virtual reality – A case study in squats and Tai Chi pushes. *Computers & Graphics* 76 (08 2018). <https://doi.org/10.1016/j.cag.2018.08.003>
 - [24] Marcelo Kallmann, Carlo Camporesi, and Jay Han. 2015. *VR-Assisted Physical Rehabilitation: Adapting to the Needs of Therapists and Patients*. Springer International Publishing, Cham, 147–168. https://doi.org/10.1007/978-3-319-17043-5_9
 - [25] Nicholas Katakis, Jonathan Tong, Oscar Ariza, Lihan Chen, Gudrun Klinker, Brigitte Röder, and Frank Steinicke. 2017. Stylo and Handfact: Modulating Haptic Perception through Visualizations for Posture Training in Augmented Reality. In *Proceedings of the 5th Symposium on Spatial User Interaction (SUI '17)*. Association for Computing Machinery, New York, NY, USA, 58–67. <https://doi.org/10.1145/3131277.3132181>
 - [26] Danielle Levac, Meghan Huber, and Dagmar Sternad. 2019. Learning and transfer of complex motor skills in virtual reality: a perspective review. *Journal of NeuroEngineering and Rehabilitation* 16 (12 2019). <https://doi.org/10.1186/s12984-019-0587-8>
 - [27] Ting-Yang Lin, Chung-Hung Hsieh, and Jiann-Der Lee. 2013. A Kinect-Based System for Physical Rehabilitation: Utilizing Tai Chi Exercises to Improve Movement Disorders in Patients with Balance Ability. In *Proceedings of the 2013 7th Asia Modelling Symposium (AMS '13)*. IEEE Computer Society, USA, 149–153. <https://doi.org/10.1109/AMS.2013.29>
 - [28] Matthew Loper, Naureen Mahmood, Javier Romero, Gerard Pons-Moll, and Michael J. Black. 2015. SMPL: A Skinned Multi-Person Linear Model. *ACM Trans. Graphics (Proc. SIGGRAPH Asia)* 34, 6 (Oct. 2015), 248:1–248:16.
 - [29] Zoe Marquardt, João Beira, Natalia Em, Isabel Paiva, and Sebastian Cox. 2012. Super Mirror: A Kinect Interface for Ballet Dancers. In *CHI '12 Extended Abstracts on Human Factors in Computing Systems (CHI EA '12)*. Association for Computing Machinery, New York, NY, USA, 1619–1624. <https://doi.org/10.1145/2212776.2223682>
 - [30] Sang Il Park and Jessica K. Hodgins. 2008. Data-Driven Modeling of Skin and Muscle Deformation. *ACM Trans. Graph.* 27, 3 (Aug. 2008), 1–6. <https://doi.org/10.1145/1360612.1360695>
 - [31] Charles Shea and Rebecca Lewthwaite. 2010. Motor skill learning and performance: A review of influential factors. *Medical education* 44 (01 2010), 75–84. <https://doi.org/10.1111/j.1365-2923.2009.03421.x>
 - [32] Rajinder Sodhi, Hrvoje Benko, and Andrew Wilson. 2012. LightGuide: Projected Visualizations for Hand Movement Guidance. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*. Association for Computing Machinery, New York, NY, USA, 179–188. <https://doi.org/10.1145/2207676.2207702>
 - [33] Maurício Sousa, João Vieira, Daniel Medeiros, Artur Arsenio, and Joaquim Jorge. 2016. SleeveAR: Augmented Reality for Rehabilitation Using Realtime Feedback. In *Proceedings of the 21st International Conference on Intelligent User Interfaces (IUI '16)*. ACM, New York, NY, USA, 175–185. <https://doi.org/10.1145/2856767.2856773>
 - [34] Arash Tadayon, Troy McDaniel, and Sethuraman Panchanathan. 2018. Functional Case Study Evaluation of the SmartGym: An Anticipatory System to Detect Body Compliance. In *Proceedings of the 3rd International Workshop on Multimedia for Personal Health and Health Care (HealthMedia'18)*. Association for Computing Machinery, New York, NY, USA, 67–71. <https://doi.org/10.1145/3264996.3265004>
 - [35] Richard Tang, Xing-Dong Yang, Scott Bateman, Joaquim Jorge, and Anthony Tang. 2015. Physio@Home: Exploring Visual Guidance and Feedback Techniques for Physiotherapy Exercises. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. Association for Computing Machinery, New York, NY, USA, 4123–4132. <https://doi.org/10.1145/2702123.2702401>
 - [36] G Tao, P S Archambault, and M F Levin. 2013. Evaluation of Kinect skeletal tracking in a virtual reality rehabilitation system for upper limb hemiparesis. In *2013 International Conference on Virtual Rehabilitation (ICVR)*. 164–165. <https://doi.org/10.1109/ICVR.2013.6662084>
 - [37] Anton Umek, Sašo Tomažič, and Anton Kos. 2015. Wearable Training System with Real-Time Biofeedback and Gesture User Interface. *Personal Ubiquitous Comput.* 19, 7 (Oct. 2015), 989–998. <https://doi.org/10.1007/s00779-015-0886-4>
 - [38] Eduardo Velloso, Andreas Bulling, and Hans Gellersen. 2013. MotionMA: Motion Modelling and Analysis by Demonstration. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. ACM, New York, NY, USA, 1309–1318. <https://doi.org/10.1145/2470654.2466171>

- [39] T von Marcard, B Rosenhahn, M J Black, and G Pons-Moll. 2017. Sparse Inertial Poser: Automatic 3D Human Pose Estimation from Sparse IMUs. *Comput. Graph. Forum* 36, 2 (2017), 349–360. <https://doi.org/10.1111/cgf.13131>
- [40] Robert Walter, Gilles Bailly, and Jörg Müller. 2013. StrikeAPose: Revealing Mid-Air Gestures on Public Displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Paris, France) (CHI '13). Association for Computing Machinery, New York, NY, USA, 841–850. <https://doi.org/10.1145/2470654.2470774>
- [41] Wenchuan Wei, Yao Lu, Eric Rhoden, and Sujit Dey. 2019. User performance evaluation and real-time guidance in cloud-based physical therapy monitoring and guidance system. *Multimedia Tools and Applications* 78, 7 (2019), 9051–9081. <https://doi.org/10.1007/s11042-017-5278-5>
- [42] Ungyeon Yang and Gerard Joungyun Kim. 2002. Implementation and Evaluation of "Just Follow Me": An Immersive, VR-based, Motion-training System. *Presence: Teleoper. Virtual Environ.* 11, 3 (2002), 304–323. <https://doi.org/10.1162/105474602317473240>