

Article

Evaluating 3D Human Motion Capture on Mobile Devices

Lara Marie Reimer ^{1,2,*} , Maximilian Kapsecker ^{1,2} , Takashi Fukushima ³  and Stephan M. Jonas ² 

¹ Department of Informatics, Technical University of Munich, Boltzmannstr. 3, 85748 Garching, Germany; max.kapsecker@tum.de

² Institute for Digital Medicine, University Hospital Bonn, Venusberg-Campus 1, 53127 Bonn, Germany; stephan.jonas@ukbonn.de

³ Department of Sports and Health Sciences, Technical University of Munich, Georg-Brauchle-Ring 60/62, 80992 München, Germany; takashi.fukushima@tum.de

* Correspondence: reimer@tum.de

Featured Application: Mobile 3D motion capture frameworks can be integrated into a variety of mobile applications. Of particular interest are applications in the sports, health, and medical sector, where they enable use cases such as tracking of specific exercises in sports or rehabilitation, or initial health assessments before medical appointments.

Abstract: Computer-vision-based frameworks enable markerless human motion capture on consumer-grade devices in real-time. They open up new possibilities for application, such as in the health and medical sector. So far, research on mobile solutions has been focused on 2-dimensional motion capture frameworks. 2D motion analysis is limited by the viewing angle of the positioned camera. New frameworks enable 3-dimensional human motion capture and can be supported through additional smartphone sensors such as LiDAR. 3D motion capture promises to overcome the limitations of 2D frameworks by considering all three movement planes independent of the camera angle. In this study, we performed a laboratory experiment with ten subjects, comparing the joint angles in eight different body-weight exercises tracked by Apple ARKit, a mobile 3D motion capture framework, against a gold-standard system for motion capture: the Vicon system. The 3D motion capture framework exposed a weighted Mean Absolute Error of $18.80^\circ \pm 12.12^\circ$ (ranging from $3.75^\circ \pm 0.99^\circ$ to $47.06^\circ \pm 5.11^\circ$ per tracked joint angle and exercise) and a Mean Spearman Rank Correlation Coefficient of 0.76 for the whole data set. The data set shows a high variance of those two metrics between the observed angles and performed exercises. The observed accuracy is influenced by the visibility of the joints and the observed motion. While the 3D motion capture framework is a promising technology that could enable several use cases in the entertainment, health, and medical area, its limitations should be considered for each potential application area.

Keywords: human motion capture; mobile motion capture; optical motion capture; consumer electronics; mHealth; dHealth



Citation: Reimer, L.M.; Kapsecker, M.; Fukushima, T.; Jonas, S.M. Evaluating 3D Human Motion Capture on Mobile Devices. *Appl. Sci.* **2022**, *12*, 4806. <https://doi.org/10.3390/app12104806>

Academic Editors: Rita M. Kiss and Alon Wolf

Received: 21 March 2022

Accepted: 6 May 2022

Published: 10 May 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Human Motion Capture (HMC) is a highly researched field and covers the detection of all kinds of human motion, including movements of the whole body or smaller parts such as the face or hands [1]. In their publications from 2001 and 2006, Moesland et al. found more than 450 publications researching vision-based HMC and analysis [1,2], not considering HMC using different technologies such as inertial or magnetic sensors.

Traditional HMC systems are bound to an off-field setting [3,4] and are expensive in installation and operation [5,6], limiting their application to professional use cases. In their review of motion capture systems in 2018, van der Kruk and Reijne identified five types of motion capture systems: Optoelectronic Measurement Systems (OMS), Inertial Sensor Measurement Systems, Electromagnetic Measurement Systems (EMS), Ultrasonic

Localization Systems (ULS), and Image Processing Systems (IPS) [7]. They introduce OMS as the gold standard for motion capture [7]. Indeed, many studies [8–13] used OMS such as the Vicon motion capture system (Vicon, Oxford, UK) [14] or the Qualisys motion capture system (Qualisys AB, Göteborg, Sweden) [15] as reference measurement systems in their studies. OMS require multiple cameras or sensors around a subject and reflection markers on the subject's anatomical landmarks, which are then captured by the cameras or sensors. The Inertial Measurement Sensor Systems rely on Inertial Measurement Units (IMU), which are placed on the subject's body to capture motion and mapped onto a rigid-body model. Examples for IMU-based systems are the Xsens systems (Xsens Technologies B.V., Enschede, The Netherlands) [16] or Perception Neuron (Noitom Ltd., Miami, FL, USA) [17]. Through the traveling time of electromagnetic or ultrasonic waves between a tagged person and a base station, EMS and ULS track the position of the subject [7,18]. In contrast to the other systems, these systems allow tracking one or more subjects' positions, but do not capture joint kinematics [7]. While the described systems are well-validated systems for HMC, their complex setup and costs prevent them from application in mHealth applications. With the advancements in technology and machine learning, IPS became more relevant in human motion capture. IPS rely on video input and different machine learning approaches to detect specific body landmarks and capture human motion. Among the most researched systems is Kinect (Microsoft Corp., Redmond, WA, USA), which uses a combination of an RGB-camera and infrared sensors and can capture motion in 3-dimensional space [10,11]. However, the Kinect still requires a specialized setup for motion capture. The offer of IPS has been extended by recent advances in technology, such as enhanced sensors and processing units. These advances enable computer-vision-based motion capture on smartphones and tablets. These IPS systems offer new possibilities for HMC in mobile scenarios such as in mHealth applications. Examples for IPS software which can run on mobile devices are OpenPose (CMU, Pittsburgh, PA, USA) [19], ARKit (Apple Inc., Cupertino, CA, USA) [20], Vision (Apple Inc., Cupertino, CA, USA) [21], and TensorFlow Pose Estimate (Google, Mountain View, CA, USA) [22]. All of these IPS can be integrated into custom applications by developers. The detection of the human body and its position is realized through computer-vision algorithms, which can use Convolutional Neural Networks (CNNs) or Part Affinity Fields (PAFs) [23]. In most systems, a predefined humanoid model is then applied to estimate the shape and kinematic structure of the tracked person [2]. The algorithms deliver the joint coordinates in two or three dimensions for every video frame.

Moeslund et al. identified three main use cases for HMC: (1) surveillance of crowds and their behavior, (2) controlling software through specific movements or gestures or controlling virtual characters in the entertainment industry such as in movies, and (3) analysis of motion for diagnostics, for example in orthopedic patients or performance improvements in athletes [2]. While use case (1) focuses on tracking multiple subjects, (2) and (3) focus on capturing body motion of a single subject and thus require tracking of several parts of the human body. Especially use case (3) offers several applications of HMC, which are often limited to professional use cases such as gait analysis [24] or sports applications [7] due to the lack of a reliable, accessible, and low-priced solution in on-field settings.

In the sports and health sector, the usage of mobile applications has significantly increased in the past years [25,26]. Research has shown that such apps can positively impact their user's health and lifestyle [27]. However, most fitness and health apps only allow limited tracking and analysis of motion [28]. While smartphone-based motion capture promises a lightweight and consumer-friendly motion capture and analysis, the software systems have only been evaluated to a limited extent. Moreover, research has been focused on 2D systems. Several studies have shown that in 2D-motion analysis, the reliability and validity of the kinematic measurements are dependent on the performed task, which reliability is measured, video quality, and position of the recording device [8,13,29,30]. Especially the camera position influences the accuracy of tracked joint angles. A slightly different viewing angle already distorts the result of the joint angle, which is why triangu-

lation with multiple devices is often performed to overcome the limitations of monocular camera setups [31]. Among mobile 2D motion capture systems, the OpenPose software is widely used and evaluated in several studies [19,23,30,32–37]. The results show that OpenPose delivers accurate biomechanical measurements, especially when tracking the joint trajectories. However, the compared joint angles differed significantly from the gold standard systems. D’Antonio et al. measured up to 9.9 degrees difference in the minima and maxima of the tracked joint angles during gait analysis [35], Nakano et al. experienced deviations of more than 40 mm in their study [37]. The measurements can be improved by using multiple devices to calculate the body position in 3D as in the study by Zago et al. [30]. Mobile 2D motion capture systems have been recently complemented by 3D motion capture algorithms, which estimate the 3D joint positions based on 2D monocular video data [20,38–42]. They detect and calculate the body’s joint coordinates in all three movement planes, making the motion capture more robust against the camera’s viewing angle. Mobile 3D motion capture frameworks could overcome the limitations of 2D motion capture systems. Some of the 3D motion capture frameworks use additional smartphone sensors such as integrated accelerometers to determine the smartphone’s position or depth sensors such as the integrated Light Detection and Ranging (LiDAR) depth sensor to additionally enhance the position detection of the human body [20,38,39]. The LiDAR data can be used to create a dense depth map from an RGB image through depth completion [43]. Among the most well-known mobile 3D motion capture systems is Apple ARKit, which released a body-tracking feature as part of their Software Development Kit (SDK) for developers in 2019 [20]. In contrast to other 3D motion capture frameworks, ARKit is free and easy to use, and widely accessible. On the latest devices, it uses the smartphone’s IMUs and integrated LiDAR sensor to improve the measurements, promising enhanced mobile motion capture. However, only a few scientific studies have evaluated the accuracy of mobile 3D motion capture frameworks and ARKit in particular. Studies mostly focused on evaluating the lower extremity tracking of ARKit [44,45].

Due to the 3D calculations, ARKit is a promising IPS software that has the potential to enable new use cases for mobile HMC previously limited to traditional HMC systems. This research evaluated ARKit’s performance against the Vicon system in a laboratory experiment in eight exercises targeting the whole body. We investigate the following two research questions:

- RQ 1: How accurate is ARKit’s human motion capture compared to the Vicon system?
- RQ 2: Which factors influence ARKit’s motion capture results?

2. Materials and Methods

2.1. Study Overview

To evaluate Apple ARKit’s body tracking accuracy, we performed a laboratory experiment in which we compared the joint angles detected ARKit against the joint angles detected by the Vicon System for marker-based, optical motion tracking. In the experiment, ten subjects were instructed to perform eight different body-weight exercises with ten repetitions each, resulting in 80 recorded exercises.

During the exercises, the complete body of the subjects was recorded using the Vicon system and two iPads running ARKit from two different perspectives. All exercises were recorded simultaneously with the Vicon system and the two iPads. The study focused on comparing the motion capture data of each iPad against the data of Vicon to answer the underlying research questions. We calculated the weighted Mean Absolute Error (wMAE) and Spearman Rank Correlation Coefficient (SRCC) between the two systems in our data analysis. In addition, we performed factor analysis using ANOVA, *t*-tests, and logistic regression to quantify the impact of specific factors on the accuracy of the ARKit performance.

2.2. Participants

We included ten subjects ($n = 10$) in the study, six males and four females. Their age ranged from 22 to 31 years, with an average of 25.7 years. The subjects' height ranged between 156 cm and 198 cm with an average of 176 cm, and their weight was between 53 kg and 90 kg, with an average of 69.5 kg. All subjects had a normal body mass index between 20.4 and 25.5 (average: 22.7) and light skin color. All subjects were in good physical condition and did not have any orthopedic or neurological impairments.

2.3. Ethical Approval and Consent to Participate

The study was conducted according to the guidelines of the Declaration of Helsinki. The ethics proposal was submitted to and approved by the Ethics Committee of the Technical University of Munich on 19 August 2021—Proposal 515/21 S. All participants were informed about the process of the study upfront, and informed written consent was obtained from all subjects involved in the study. Due to the non-interventional character of this study, the risks involved for the study participants were low. We further minimized the risk through a sports scientist who supervised the physiologically correct execution of all exercises during the study, preventing the participants from performing potentially harmful movements.

2.4. Exercise Selection

Eight exercises were selected: Squat, Front Lunge, Side Squat, Single Leg Deadlift, Lateral Arm Raise, Reverse Fly, Jumping Jacks, and Leg Extension Crunch. The main objective of the exercise selection was to create a full-body workout to track all selected joints from different angles.

All exercises were tested for the suitability of tracking in both systems to ensure stable tracking of the angles. Both ARKit and the Vicon system exposed problems with the correct detection of exercises, where more extensive parts of the body were hidden from the cameras, for example, push-ups, and were therefore excluded. The testing was done in two steps: (1) We manually inspected the screen recording to see if the ARKit app model recognized the subject. (2) We checked the screen recording to whether the ARKit model overlaid with the subject's body parts during all parts of the exercise and whether the Vicon system could track all markers in the majority of recorded frames so that the full joint trajectory could be calculated.

Only if both requirements were fulfilled, we selected the exercise for the study. The final exercise selection included eight exercises. Their execution (E, see Figure 1) and targeting muscle groups (TMG) are explained in the following, and tracked joint angles (TJA) are explained in the following.

(I) Squat: (E:) The subject starts this exercise in an upright standing position. The subject squats down from the starting position by flexing the ankle, knee, and hip without movement compensations such as flexing the trunk and raising the heel. Each subject was asked to hold their arms stretched in front of the body. (TMG:) This exercise targets the lower body, especially the gluteus, quadriceps, hamstrings, and calves. (TJA:) The tracked joint angles include the left and right hip, and left and right knee.

(II) Front Lunge: (E:) The starting position of the exercise is an upright standing with spreading legs front and back. The arms' position is the same as the squat. From the starting position, the subject goes down by flexing the ankle, knee, and hip in the front leg, flexing the knee and hip, and raising the heel in the back leg. (TMG:) This exercise targets lower body muscles, especially the gluteus, quadriceps, hamstrings, and calves. (TJA:) The tracked joint angles include the left and right hip, and left and right knee.

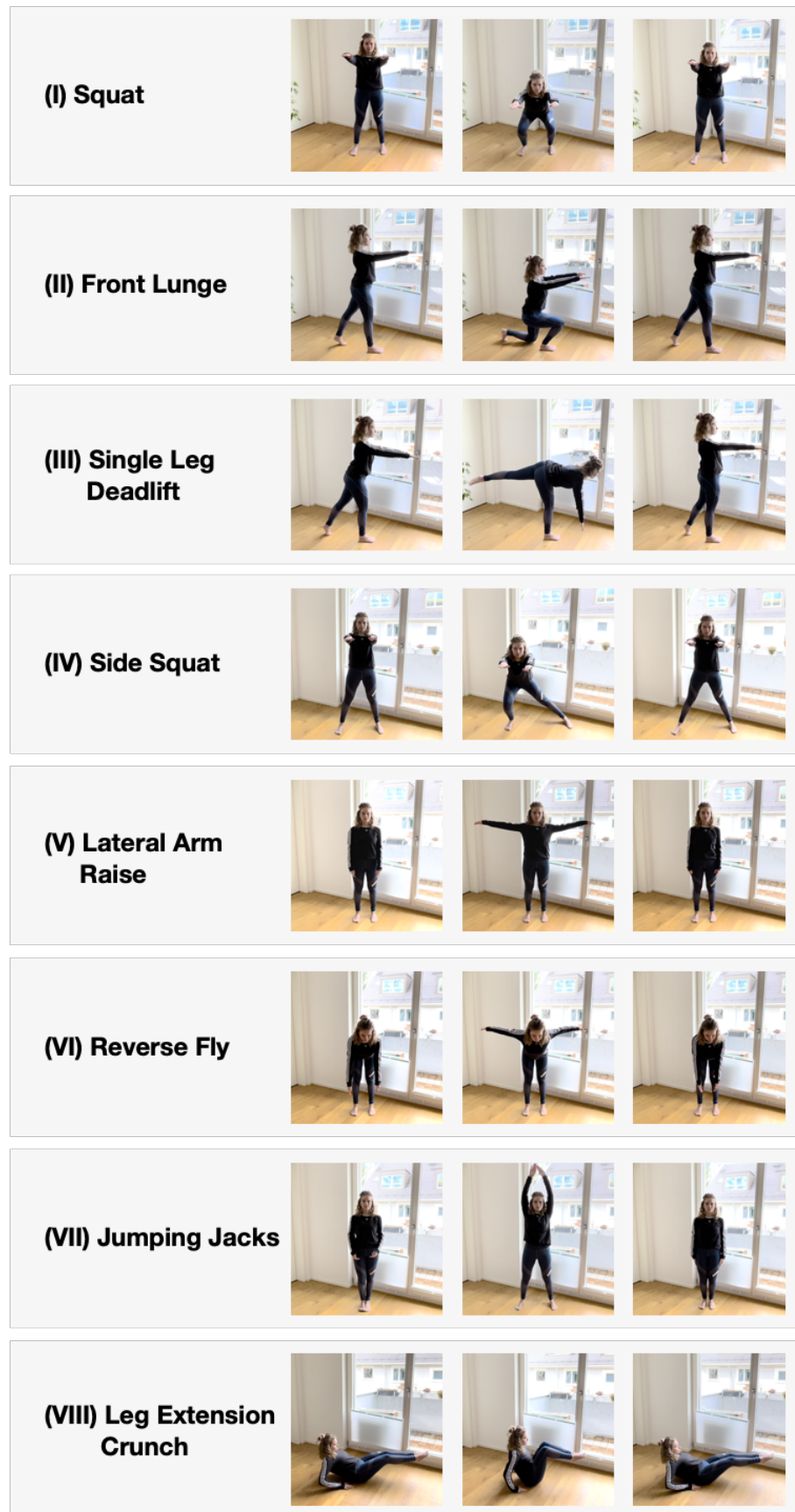


Figure 1. The execution of all eight exercises as seen from the frontally positioned iPad. The body orientation was chosen to maximize the visible parts of the body.

(III) Side Squat: (E:) The starting position of the exercise is an upright standing with spreading legs laterally. The arms' position is the same as the squat. From the starting point, the subject squats down with either side with either leg while the other leg is kept straight. (TMG:) This exercise targets similar muscle groups to squats, focusing on adductor muscles. (TJA:) The tracked joint angles include the left and right hip, and left and right knee.

(IV) Single Leg Deadlift: (E:) The starting position of the exercise is an upright standing with a single leg. The arms' initial position is the same as in the Squat. The subject leans forward from the starting position by flexing the hip with minimum knee flexion. As the subject leans forward, the arms should be hung in the air. The other side of the leg in the air should be extended backward to maintain balance as the subject leans forward. (TMG:) The exercise targets lower body muscles, especially the hamstring and gluteal muscles. (TJA:) The tracked joint angles include the left and right hip, and left and right knee.

(V) Lateral Arm Raise: (E:) The subject starts the exercise in an upright standing position. Then, the subject laterally abducts the arms. (TMG:) The exercise targets upper body muscles, especially the deltoid muscles. (TJA:) The tracked joint angles include the left and right shoulder, and left and right elbow.

(VI) Reverse Fly: (E:) The subject leans forward with slight knee flexion and hangs the arms in the air in a starting position. The subject horizontally abducts the arms from the position without raising the upper body. (TMG:) The exercise targets upper body muscles such as the rhomboid, posterior deltoid, posterior rotator cuff, and trapezius muscles. (TJA:) The tracked joint angles include the left and right shoulder, and left and right elbow.

(VII) Jumping Jack: (E:) This exercise starts from an upright standing position. Then, the subject abducts both sides of the legs and arms simultaneously with a hop. (TMG:) This exercise targets lower body and upper body muscles, especially the gluteal and deltoid muscles. (TJA:) The tracked joint angles include the left and right shoulder, left and right elbow, left and right hip, and left and right knee.

(VIII) Leg Extension Crunch: (E:) The subject starts this exercise by sitting down on the ground with a backward lean of the upper body. The subject should place the hands on the ground to support the upper body as leaning backward. Then, the subject brings the legs in the air with knee and hip flexion. From the position, the subject extends the knee and hip horizontally on both sides together. (TMG:) This exercise targets core muscles, especially abdominal muscles. (TJA:) The tracked joint angles include the left and right hip, and left and right knee.

2.5. Data Collection

We prepared the laboratory before the subjects arrived to ensure similar conditions for all recordings. Four tripods were positioned, each of them approximately three meters away from the area of the subjects' position to enable tracking of the entire body. Two tripods held an iPad Pro 11" (2021 Model; Apple Inc., Cupertino, CA, USA), which were used to run the ARKit motion capture. Two other tripods were equipped with regular cameras to record videos of the experiment. One iPad and one camera were placed facing the subject's position frontally, the other iPad and camera were placed at an approximate angle of 30° facing the subject, as shown in Figure 2. The Vicon system (Nexus 2.8.2, Version 2.0; Vicon Motion Systems Ltd., Oxford, UK) was installed on the lab ceiling and configured to track the subjects' whole body.

We developed a protocol to guarantee a similar experiment execution for all participants. The experiment consisted of three phases: (1) the onboarding, (2) the explanation of the exercises, and (3) performing the exercises. During phase (1), the participants entered the lab. We explained the setup, and the participants signed the consent forms. In phase (2), a sports scientist explained each of the eight exercises and showed the participants how they are performed. The participants were asked to perform the exercises once under the supervision of the sports scientist to guarantee correct execution. The actual experiment was performed in phase (3). The participants performed ten repetitions of each exercise.

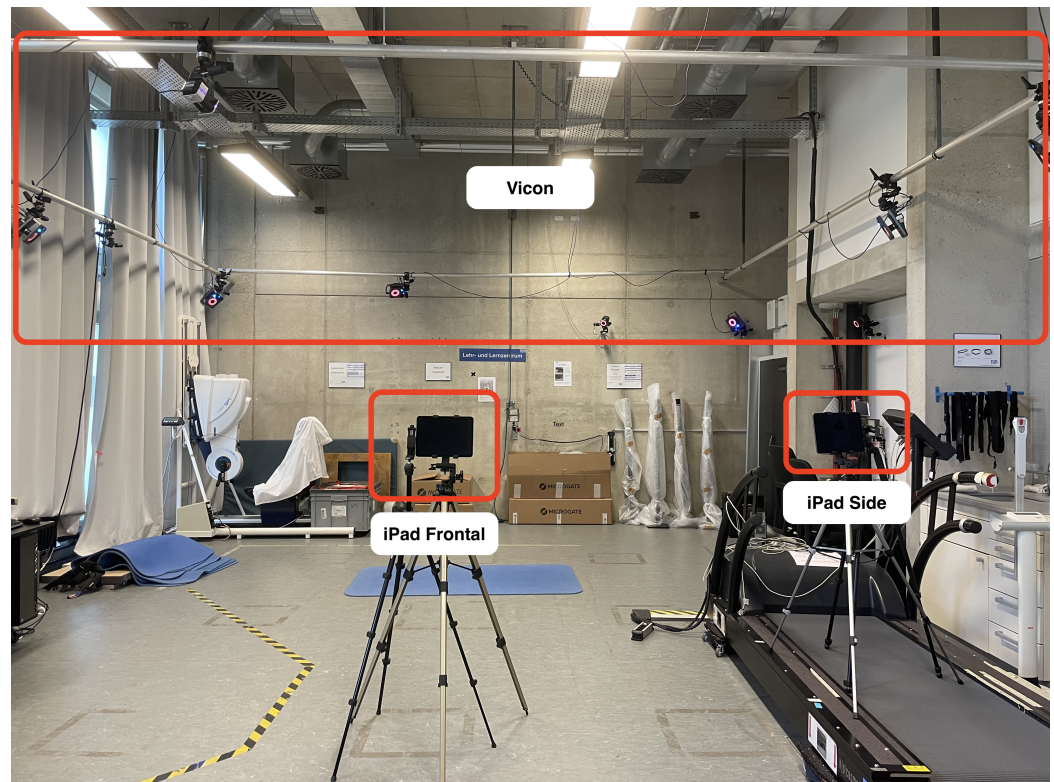


Figure 2. The experiment setup, showing the positioning of the recording devices and the subject.

2.5.1. Vicon Setup

The Vicon setup consisted of 14 infrared cameras. The setup included eight MX-T10-S cameras, four Vero v2.2 cameras, and two Bonita 10 cameras. All cameras were set to a sampling frequency of 250 Hz. We used the Nexus software (version 2.8.2) with the Full-Body Plug-in Gait marker placement model provided by Vicon Motion Systems, Ltd. [46] to capture the motion. A Vicon calibration wand was used to calibrate all the Vicon cameras and determine the coordinate system. Static calibration was done by capturing a subject performing a T-pose.

2.5.2. ARKit Setup

The ARKit setup included two iPad Pro 11" 2021 with an M1 processor and an additional LiDAR sensor for depth information. Both iPads ran a custom-developed software based on the ARKit 5 framework provided by Apple Inc., which was used for extracting the motion capture information from the iPads' sensors. Both iPads recorded the motion capture data independently and were not synchronized. The motion capture data included the timestamp of the detection, the performed exercise, and the three-dimensional, positional information of 14 body joints. These data were later used to calculate the joint angles. All joint coordinates are given relatively to the pelvis center, which serves as the origin of ARKit's coordinate system. ARKit differentiates between bigger joints, which are actively tracked, and calculated joints, which are smaller joints such as the toes and fingers. We decided only to include actively tracked joints in our comparison, as previous tests showed that the calculated position of the smaller joints and their related angles rarely change. The ARKit data were recorded with a default sampling frequency of 60 Hz. However, the sampling frequency of ARKit is variable, as ARKit internally only updates the joint positions when a change is detected. This means that if a subject is standing still, fewer data points are received from ARKit and more when the subject is moving fast. As the toe and finger joints are calculated by ARKit and not actively recognized, we limited the comparison to the actively tracked joints: shoulders, elbows, hips, and knees.

2.5.3. Data Export

After each recorded subject, the collected motion data were exported from the three systems: the frontally positioned iPad (iPad Frontal), the iPad set in a 30° Side Angle (iPad Side) (Figure 2), and the Vicon system. The motion data were stored in CSV files and included the joint center coordinates for each detected frame for the three systems separately. The ARKit data were exported in one file per iPad, resulting in two CSV files per subject. For the Vicon system, each exercise was stored in a separate CSV file. In addition, an XCP file was exported from the Vicon system, which contained meta-information about the cameras, including the start and end timestamps of each recording.

Due to export problems, the upper body joint coordinates of the iPad Side were only included for three of the ten subjects. The Vicon system could not track each joint coordinate throughout the whole exercise due to hidden markers, leading to gaps in the exported data. Smaller gaps were compensated during the Data Analysis, whereas more significant gaps led to the exclusion of the respective angle.

2.6. Preprocessing & Data Analysis

The basis for the data analysis part is 220 files, 22 for each subject. It contains two comma-separated value (CSV) files from the respective ARKit systems (frontal and side view) and ten CSV files from the Vicon system, which records each exercise in a separate file. The remaining ten files are given in the XCP format, which contains the relevant metadata of the Vicon system, such as camera position, the start time, and the end time of the data acquisition process. The following preprocessing steps are performed for each subject to merge all files into a data frame for further analysis.

The Vicon and ARKit data are modified to fit a matrix-like structure in which the rows represent time and columns the joints. Augmentation enhances the data with information such as the timestamp, subject, exercise, and in the case of ARKit, whether the values were recorded frontal or lateral.

The Sections 2.5.1 and 2.5.2 explain different sampling rates for the systems and the non-equidistant sampling rate of ARKit (57 Hz on average). It motivates to evaluate strategies to merge the system's data based on the timestamp. Vicon samples the data at a frequency of 250 Hz and implies a maximum of 2 ms distance for a randomly chosen timestamp. Due to this maximal possible deviation, the nearest timestamp is the criterion for merging the Vicon data onto the ARKit data.

The Vicon system records absolute coordinates, while the ARKit system provides normalized coordinates relative to the center of the hip. It still allows for comparing angles since they are invariant under scaling, rotating, translating, and reflecting the coordinate system. Accordingly, the adjacent three-dimensional joint coordinates extraction calculates the angles of interest (AOI). An angle θ is determined by three joints $A, B, C \in \mathbb{R}^3$ or associated vectors $\vec{v}_1 = A - B$ and $\vec{v}_2 = C - B$ given the formula

$$\theta = \arccos \frac{v_1 \cdot v_2}{\|v_1\|_2 \|v_2\|_2}$$

The data reveal a time lag which leads to a misalignment between the Vicon and ARKit angles along the time axis. Accordingly, the related time series require shifting with the objective to maximize the mutual Pearson correlation coefficient. The shift operation is subjected to a maximum of 60 frames to each side. It includes the assumption that the time series of the two systems match best if they exhibit similar behavior in their linear trends. Figure 3 shows two examples of misaligned time series on the left and the result of the shift on the right. The time series alignment is performed brute force and individually for any combination of view, subject, exercise, and AOI. The procedure outputs 1048 ARKit-Vicon time series pairs, 634 for the comparison Vicon—iPad Frontal, and 414 for the comparison Vicon—iPad Side. The number does not correspond to $2 \times 10 \times 8 \times 8 = 1280$ pairs due to the missing ARKit recordings of the upper body joints for lateral recording.

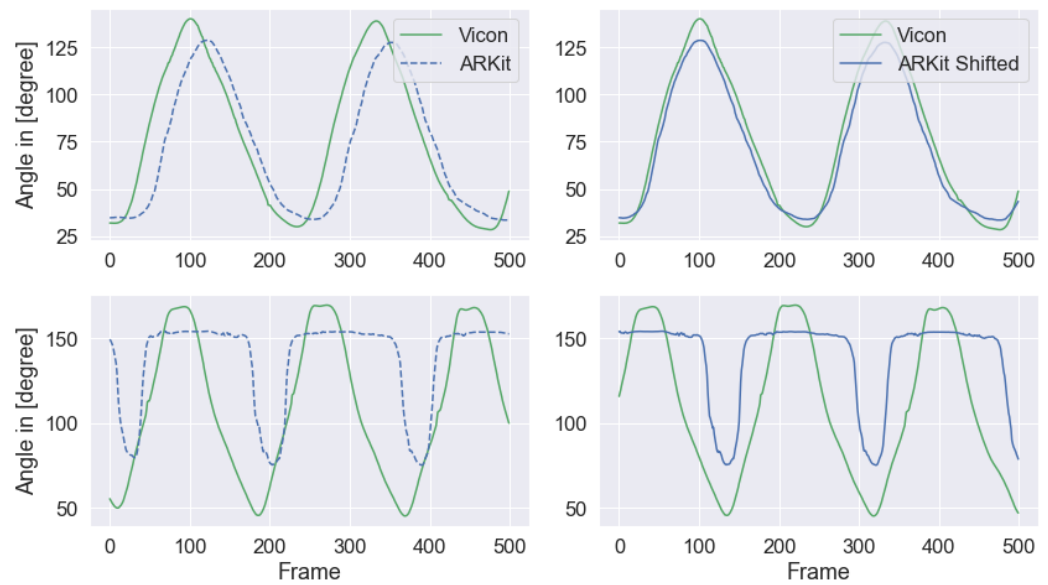


Figure 3. Shift of the data.

Computing two metrics validates the angle similarity of the systems for each pair of time series, the mean absolute error (MAE) and the non-parametric Spearman’s rank correlation coefficient (SRCC). The obtained MAE and SRCC values of the 1048 time series are aggregated according to predefined grouping criteria, such as exercise, angle, or view. Calculating the sample size’s weighted mean and standard deviation (std) defines a grouping operation for the MAE (Table 1). SRCC values require first a transformation to a normally distributed random variable using the Fisher z-transformation

$$z = \frac{1}{2} \ln\left(\frac{1+r}{1-r}\right) \tag{1}$$

where r is the SRCC. It constitutes the prerequisite to applying the averaging operation along with the variables. The result is again a normally distributed variable that needs back transformation into the correlation space using the inverse of (1).

Table 1. The aggregated wMAE values for all joint angles.

Angle	wMAE
leftElbow	24.0° ± 17.43°
leftHip	16.91° ± 10.67°
leftKnee	16.61° ± 7.47°
leftShoulder	20.01° ± 14.89°
rightElbow	20.0° ± 15.32°
rightHip	20.17° ± 11.25°
rightKnee	17.57° ± 7.25°
rightShoulder	17.39° ± 12.18°

A drawback of the MAE is the lack of interpretation regarding systematic over- or underestimation of the angles. The mean error (ME), which is the average of the time series pair’s difference, can conclude the occurrence of bias but at a granular level, for example segments of the exercise. However, aggregation of the ME is prone to involve effects such as error cancellation. The ratio of ME and MAE, for instance $\frac{ME}{MAE}$, draws insights into the occurrence of systematic bias (Figure A4). A value close to ±1 implies less tendency of ARKit to fluctuate around the Vicon’s angle estimation, for example either under-, perfect- or overestimation takes place. Values nearby zero indicate the ME’s cancellation effect

(over- and underestimation) but require further analysis, such as the difference between MAE and ME, for conclusions.

One-way analysis of variance (ANOVA) is performed to quantify the effects of the categorical metadata such as angle (fixed effect), exercise (fixed effect), and subject (random effect), on the continuous variable MAE. The random effect was taken into account performing one-way ANOVA using a random effects model. The distribution of MAE shows a divergence towards the normal distribution, which is one of the requirements in ANOVA. However, research verified robustness in violating this assumption in certain bounds [47]. A logarithm (basis 10) transform on the MAE variable ensures stronger normalization (Appendix A, Figure A1). In particular, it makes the model multiplicative and more robust to dispersion. The visual inspection of histograms reveals a lack of homogeneous intergroup variance and motivates to apply Welch's ANOVA. Finally, the Games-Howell post-hoc test [48] compares the individual categorical factors for significant results (here defined as an effect size larger than 0.1).

Besides view (frontal or side), the binary independent variables are the body segment of the angle (lower or upper) and information on the movement of the pelvis. The latter is declared as the variable center moved and indicates whether the proper execution of the exercise involves the movement of the pelvis's center, the origin of the ARKit coordinate system. To quantify the binary variables' effect, we fitted a logistic regression model based on the MAE and applied Welch's *t*-test. The results, including β coefficient, R^2 *p*-value, and confidence interval, are compiled in a table.

Assumptions about the data are made and can restrict the interpretation of the results. A more detailed outline of this topic is given in the limitations section (Section 5.8).

3. Results

3.1. Weighted Mean Absolute Error

3.1.1. Aggregated Results

The data analysis exposed a wMAE of $18.80^\circ \pm 12.12^\circ$ degrees for all angles in the whole data set. The wMAE across all exercises, views, and angles is visualized in Figure 4 to enable more profound insights into the performance based on exercises and joint angles.

The data exposed high differences in the detected error rates with the wMAE ranging between $3.75^\circ \pm 0.99^\circ$ (Lateral Arm Raise, Left Elbow, Side) and $47.06^\circ \pm 5.11^\circ$ (Side Squat, Left Elbow, Side), depending on the performed exercise and observed joint. To generate better insights into the different factors, we aggregated the wMAE by angle, performed exercise, view, and subject.

Considering the aggregated wMAE for the individual joints (Table 1), the mean value ranged between $16.61^\circ \pm 7.47^\circ$ for the left knee up to $24.00^\circ \pm 17.43^\circ$ for the left elbow. The left hip exposed a wMAE of $16.91^\circ \pm 10.67^\circ$, followed by the right shoulder with a wMAE of $17.39^\circ \pm 12.18^\circ$ and the right knee with a value of $17.57^\circ \pm 7.25^\circ$. The right elbow had a wMAE of $20.00^\circ \pm 15.32^\circ$, the left shoulder $20.01^\circ \pm 14.89^\circ$ and the right hip $20.17^\circ \pm 11.25^\circ$.

The observed wMAE differed between the exercises, with the Lateral Arm Raise ($9.56^\circ \pm 6.13^\circ$), Jumping Jacks ($10.09^\circ \pm 3.81^\circ$), Single Leg Deadlift ($11.35^\circ \pm 5.04^\circ$), Reverse Fly ($15.80^\circ \pm 8.5^\circ$), Leg Extension Crunch ($18.15^\circ \pm 8.21^\circ$), and Front Lunge ($18.19^\circ \pm 8.98^\circ$) exposing significantly lower error rates than the Side Squat ($30.49^\circ \pm 12.73^\circ$) and the Squat ($33.79^\circ \pm 10.25^\circ$) (Table 2).

When only considering the targeted joints, the wMAE ranged between $3.75^\circ \pm 0.99^\circ$ (Lateral Arm Raise, Left Elbow, Side View) and $38.41^\circ \pm 6.66^\circ$ (Squat, Right Hip, Frontal View). The exercises Lateral Arm Raise, Reverse Fly, and Single Leg Deadlift performed best with wMAE values below 15.00° in the relevant joints for the respective exercises. The wMAE of Jumping Jacks, Front Lunge, and Leg Extension Crunch remained below 25.00° across the targeted joints. The Squat and Side Squat Exercises exposed error rates of up to 38.41° in the targeted joints and thus performed worst in the experiment.

Exercise	Angle View	leftElbow	rightElbow	leftShoulder	rightShoulder	leftHip	rightHip	leftKnee	rightKnee
Front Lunge	Frontal	35.26 ± 8.06	23.75 ± 7.44	8.39 ± 4.46	12.36 ± 3.23	10.83 ± 4.12	20.14 ± 6.82	13.99 ± 2.76	13.97 ± 3.90
	Side	36.66 ± 2.49	33.68 ± 4.20	14.7 ± 0.62	17.6 ± 2.51	12.54 ± 3.12	22.26 ± 3.19	17.57 ± 4.07	17.96 ± 2.93
Jumping Jacks	Frontal	7.66 ± 3.01	7.02 ± 2.50	6.60 ± 1.27	7.58 ± 1.50	8.4 ± 1.57	8.33 ± 1.86	15.15 ± 2.56	15.02 ± 2.00
	Side	9.14 ± 1.07	9.32 ± 0.30	6.49 ± 1.18	7.77 ± 0.42	9.39 ± 1.97	9.93 ± 2.58	14.9 ± 2.12	13.94 ± 1.80
Lateral Arm Raise	Frontal	7.51 ± 3.35	7.36 ± 2.67	6.65 ± 1.34	6.81 ± 1.97	5.01 ± 2.67	4.80 ± 2.7	17.67 ± 4.17	17.61 ± 3.32
	Side	3.75 ± 0.99	5.43 ± 1.36	5.5 ± 0.39	5.69 ± 0.42	4.93 ± 2.14	6.84 ± 3.28	17.76 ± 4.39	17.44 ± 3.86
Leg Extension Crunch	Frontal	18.94 ± 6.77	19.49 ± 5.99	31.98 ± 10.20	17.39 ± 5.80	10.86 ± 2.96	11.49 ± 3.86	14.81 ± 4.18	16.67 ± 3.07
	Side	19.78 ± 6.42	26.07 ± 10.73	36.33 ± 4.94	21.29 ± 4.66	14.0 ± 4.81	16.44 ± 6.24	17.18 ± 3.88	20.17 ± 6.00
Reverse Fly	Frontal	8.90 ± 3.91	10.27 ± 3.98	11.84 ± 4.02	10.63 ± 3.80	25.08 ± 6.27	27.73 ± 6.31	15.54 ± 7.54	15.01 ± 7.23
	Side	7.89 ± 2.10	14.85 ± 5.22	14.72 ± 3.68	8.69 ± 2.99	22.32 ± 7.14	20.98 ± 7.48	15.37 ± 8.10	11.98 ± 7.05
Side Squat	Frontal	46.73 ± 14.20	42.48 ± 14.29	41.82 ± 9.72	36.21 ± 8.84	22.62 ± 7.63	36.41 ± 5.21	16.50 ± 6.13	26.48 ± 4.13
	Side	47.06 ± 5.11	27.74 ± 4.13	30.04 ± 2.17	23.67 ± 1.60	26.19 ± 8.28	30.25 ± 3.95	16.17 ± 5.88	21.54 ± 2.76
Single Leg Deadlift	Frontal	21.07 ± 4.82	8.59 ± 4.60	11.78 ± 2.77	10.82 ± 2.96	10.57 ± 3.33	14.37 ± 4.20	8.66 ± 2.47	9.25 ± 3.49
	Side	13.26 ± 1.06	6.11 ± 0.98	9.39 ± 1.36	7.13 ± 1.59	12.28 ± 4.96	14.91 ± 4.61	8.17 ± 2.85	10.0 ± 3.40
Squat	Frontal	44.35 ± 17.93	37.85 ± 16.51	39.23 ± 9.04	37.49 ± 8.70	35.37 ± 6.46	37.41 ± 6.66	29.94 ± 4.42	30.36 ± 3.72
	Side	45.40 ± 10.51	31.86 ± 9.57	36.56 ± 5.05	27.77 ± 3.06	32.04 ± 4.63	30.05 ± 4.84	27.98 ± 3.31	23.73 ± 2.00

Figure 4. Pivot Table of the weighted Mean Absolute Error (wMAE) in degrees distributed over the eight exercises and the eight tracked angles, each measured from the two iPad perspectives *Frontal* and *Side*. The dashed boxes indicate which joints were specifically targeted by the respective exercise. The heatmap visualizes the performance of the individual joints per exercise, with darker green color referring to a lower error rate and darker orange color referring to higher error rates.

When only considering the targeted joints per exercise, the wMAE was reduced for all exercises except the Jumping Jacks, where the wMAE remained the same (Table 2).

Table 2. The wMAE values for all exercises when considering all angles and only the targeted angles per exercise.

	All Angles	Targeted Angles
Front Lunge	18.19° ± 8.98°	16.17° ± 5.48°
Jumping Jacks	10.09° ± 3.81°	10.09° ± 3.81°
Lateral Arm Raise	9.56° ± 6.13°	6.66° ± 2.41°
Leg Extension Crunch	18.15° ± 8.21°	15.14° ± 5.34°
Reverse Fly	15.80° ± 8.5°	10.67° ± 4.31°
Side Squat	30.49° ± 12.73°	24.56° ± 8.63°
Single Leg Deadlift	11.35° ± 5.04°	10.91° ± 4.41°
Squat	33.79° ± 10.25°	30.93° ± 6.19°

The difference between the view of the recording device was smaller than the observed differences between the exercises, with an wMAE of 17.91° ± 9.68° for the side view and 19.35° ± 13.38° for the frontal view.

When considering the different subjects, the observed wMAE was relatively consistent among the individuals, with mean values ranging from 16.20° ± 9.44° to 22.32° ± 17.08°.

3.1.2. Bias of the ARKit System

For detecting a possible bias of over- and underestimation of the ARKit data, we investigated the ME and the ratio of ME/MAE. The aggregated results of the ME/MAE ratio exhibits only seven values below 0.1 for the exercise—angle—view configurations (Appendix B Figure A3 for the ME, Appendix B Figure A4 for ratio ME/MAE). In 4 of these cases, the wMAE is above 10°: Front Lunge—left hip—Frontal, Jumping Jacks—left knee—Frontal, Jumping Jacks—right knee—Frontal, and Leg Extension Crunch—left elbow—Frontal. Most other values remain relatively close to 1 or −1.

3.2. Spearman Rank Correlation

The whole dataset exposed a mean Spearman Rank Correlation Coefficient of 0.76. The *p*-value was below 0.01 for 1019 of the 1048 exercises. A detailed overview of the individual SRCCs, including the standard deviation for the exercises, is visualized in Figure 5.

Exercise	Angle View	leftElbow	rightElbow	leftShoulder	rightShoulder	leftHip	rightHip	leftKnee	rightKnee
Front Lunge	Frontal	0.22	0.16	0.67	0.65	0.49	0.93	0.91	0.95
	Side	-0.13	-0.20	0.59	0.62	0.63	0.97	0.92	0.97
Jumping Jacks	Frontal	0.36	0.25	0.91	0.90	0.43	0.42	0.32	0.40
	Side	0.63	0.47	0.93	0.91	0.32	0.66	0.43	0.70
Lateral Arm Raise	Frontal	0.79	0.82	0.96	0.96	0.54	0.25	0.22	0.13
	Side	0.78	0.68	0.99	0.96	0.61	0.45	0.26	0.17
Leg Extension Crunch	Frontal	0.55	0.69	0.44	0.85	0.94	0.92	0.92	0.90
	Side	0.26	0.68	0.32	0.81	0.90	0.85	0.93	0.89
Reverse Fly	Frontal	0.45	0.47	0.87	0.84	0.80	0.79	0.50	0.53
	Side	0.45	0.33	0.80	0.84	0.74	0.76	0.43	0.49
Side Squat	Frontal	-0.05	-0.08	0.56	0.65	0.90	0.90	0.63	0.93
	Side	0.13	-0.03	0.25	0.52	0.91	0.98	0.63	0.97
Single Leg Deadlift	Frontal	0.33	0.69	0.86	0.89	0.95	0.77	0.74	0.51
	Side	0.55	0.63	0.97	0.97	0.94	0.70	0.78	0.49
Squat	Frontal	-0.06	-0.15	0.71	0.70	0.75	0.79	0.84	0.89
	Side	-0.27	0.05	0.77	0.60	0.88	0.95	0.90	0.97

Figure 5. Pivot Table of the average Spearman Rank Correlation Coefficients (SRCC) distributed over the eight exercises and the eight tracked angles, each measured from the two iPad perspectives *Frontal* and *Side*. The dashed boxes indicate which joints were specifically targeted by the respective exercise. The heatmap visualizes the performance of the individual joints per exercise, with darker green color referring to a higher positive correlation and darker orange color referring to a higher negative correlation.

The SRCC varied between the tracked angles with a range of -0.27 to 0.99 as mean values per exercise and angle as displayed in Figure 5. When considering the results aggregated per joint angles (Table 3), all negative correlations were observed for the elbow angles (left elbow 0.36 , right elbow 0.42) in both iPad views, with the side view performing worse than the frontal view. The shoulder angles exposed a mean SRCC of 0.81 for both shoulders. Knee and hip joints were also tracked with moderate SRCC values (left hip: 0.82 , right hip: 0.84 , left Knee: 0.75 , right knee: 0.81).

Table 3. The aggregated SRCC values for all joint angles.

Angle	SRCC
leftElbow	0.36
leftHip	0.82
leftKnee	0.75
leftShoulder	0.81
rightElbow	0.42
rightHip	0.84
rightKnee	0.81
rightShoulder	0.81

While the SRCCs differed between the exercises, all of them exposed moderate linear correlations with values above 0.5 (Table 4). The Leg Extension Crunch showed a correlation of 0.84 . Front Lunge correlated with 0.80 , followed by the Single Leg Deadlift with an SRCC of 0.79 . The Squat and Side Squat exercises showed a correlation of 0.78 . The SRCC of the

Lateral Arm Raise was 0.68, and the SRCC of the Reverse Fly was 0.67. The Jumping Jacks performed worst with a correlation of 0.60.

Similar to the wMAE, considering only the relevant joints for the specific exercises positively influenced the SRCCs of all exercises except for the Jumping Jacks, where it remained the same, and the Single Leg Deadlift, where it was reduced by 0.01 (Table 4).

Table 4. The average SRCC values for all exercises when considering all angles and only the targeted angles per exercise.

	SRCC All Angles	SRCC Targeted Angles Only
Front Lunge	0.80	0.91
Jumping Jacks	0.60	0.60
Lateral Arm Raise	0.68	0.91
Leg Extension Crunch	0.84	0.91
Reverse Fly	0.67	0.69
Side Squat	0.78	0.91
Single Leg Deadlift	0.79	0.78
Squat	0.78	0.89

Comparing the two positions of the iPads, the side view performed slightly better than the frontal view, with SRCCs of 0.80 and 0.73, respectively.

Similar to the wMAE, the SRCC is relatively consistent across the recorded subject, with values between 0.72 and 0.82.

3.3. Factor Analysis

3.3.1. ANOVA Analysis

To further investigate the influence of the observed exercise, angle, and subject on the performance of ARKit, we performed a Welch ANOVA factor analysis on the Mean Absolute Error for the factors *Exercise* and *Angle* and a random effects model for the factor *Subject*. The MAE exhibited a high dependency on the observed exercise with an effect size of $\eta^2 = 0.51$ ($p = 0.00$). It did not expose a dependency on the observed angle ($\eta^2 = 0.03$, $p = 0.00$). The random effects model analysis did not exhibit an influence of the subject, with 0.29% of the variance explained by the subject (Table 5).

Table 5. The results of the Random Effects ANOVA.

Random Effects			
Groups	Name	Variance	Std. Dev.
Subject	(Intercept)	0.001312	0.03622
Residual		0.458310	0.67699
Fixed Effects			
	Estimate	Std. Error	t value
(Intercept)	2.70803	0.02389	113.4

To further investigate the influencing factors of the performed exercise in the MAE, we performed a Post-hoc analysis using the Games-Howell test (Appendix C, Table A1). The exercise analysis exhibits significant differences between 20 of the 28 exercise pairs.

3.3.2. Welch t-Test Analysis

All binary influencing factors of the MAE were analyzed using Welch’s *t*-test (Table 6). The results of the *t*-test showed a dependency on the pelvic center movement ($cohen - d = 0.82$, $power = 1.00$, $p = 0.00$). No dependency was measured for the view ($cohen - d = 0.01$, $power = 0.06$, $p = 0.82$), and whether the measured angle is a lower body angle ($cohen - d = 0.01$, $power = 0.05$, $p = 0.88$).

Table 6. The results of the Welch *t*-test Analysis.

T	dof	Alternative	p-Value	CI95%	Cohen-d	BF10	Power	Response	Categorical
-0.22	966.81	two-sided	0.82	[-0.09, 0.07]	0.01	0.073	0.06	LogMAE	View
-0.15	725.74	two-sided	0.88	[-0.1, 0.08]	0.01	0.072	0.05	LogMAE	LowerBody
-13.20	1045.97	two-sided	0.00	[-0.59, -0.44]	0.82	3.266×10^{33}	1.00	LogMAE	CenterMoved

3.3.3. Logistic Regression Analysis

In addition to the *t*-test, we applied logistic regression to the three variables *View*, *LowerBody*, and *CenterMoved* (Table 7). The logistic regression model for the *LowerBody* shows a slight effect with $\beta - coef = 0.0684$ ($p = 0.00$). The model exposed a *Pseudo - R²* of 0.165. While the *View* model exposed no significant effect ($\beta - coef = 0.0141$, $p = 0.00$), the fitness of the model is low (*Pseudo - R²* = -0.019). The *CenterMoved* variable showed no effect ($\beta - coef = 0.0018$, $p = 0.575$). Similar to the *View* variable, the *Pseudo - R²* of 0.000 indicated bad fitness of the model to explain the data.

Table 7. The results of the logistic regression.

Variable	β -coef	std	z	P > z	[0.025	0.975]	<i>Pseudo-R²</i>
View	0.0141	0.003	4.329	0.000	0.008	0.020	-0.019
Lower Body	0.0684	0.005	13.374	0.000	0.058	0.078	0.165
Center Moved	0.0018	0.003	0.561	0.575	-0.004	0.008	0.000

4. Findings

While the results showed that ARKit is generally capable of tracking human body motion, the accuracy of the joint angles is highly variable and dependent on several factors, especially the performed exercise.

4.1. RQ 1: How Accurate Is ARKit’s Human Motion Capture Compared to the Vicon System?

To answer RQ 1, we investigated both the wMAE and the SRCC of the experiment data. A wMAE of 0° and an SRCC of 1.0 would represent a perfect accuracy of ARKit’s human motion capture. The ARKit data showed a MAE of 18.80° and an average SRCC of 0.76 for the whole data set, with variations when examining different joints and exercises. Based on the results of the ANOVA analysis, the accuracy mainly depends on the observed angle and exercise. However, the accuracy could be influenced by other additional factors which were not specifically targeted by the performed experiment. Remarkably, ARKit was able to achieve an almost perfect correlation and accuracy for some exercise executions in specific angles (Figure 6). In many cases, the movement pattern is recognizable in the ARKit data. Still, the amplitude is reduced, or a baseline drift on the *y*-axis is observable (Figure 7, which explains the good correlation but relatively high wMAE values. In some cases, the ARKit data exhibits high wMAE values and no or even a negative SRCC. These effects often occurred in the elbow joints, especially when the lower body joints moved and the upper body joints were held straight, such as in the Squat or Side Squat exercises. In this situation, ARKit often failed at detecting the movement correctly (Figure 8), which is visible both in the high wMAE and the low to negative correlation values for the elbow angles. In general, the accuracy was lower in those exercises where the root position did not remain stable, including the Front Lunge, Side Squat, and Squat exercises. The results of the factor analysis further confirmed these results.

To investigate whether a systematic baseline drift can be observed in the ARKit data, we aligned the ARKit and Vicon data via cross-correlation. We measured the *y*-axis offset (Figure 9). As the offset was normally distributed around 0, no systematic baseline drift was present in the recorded data set, indicating that other factors cause shifts.

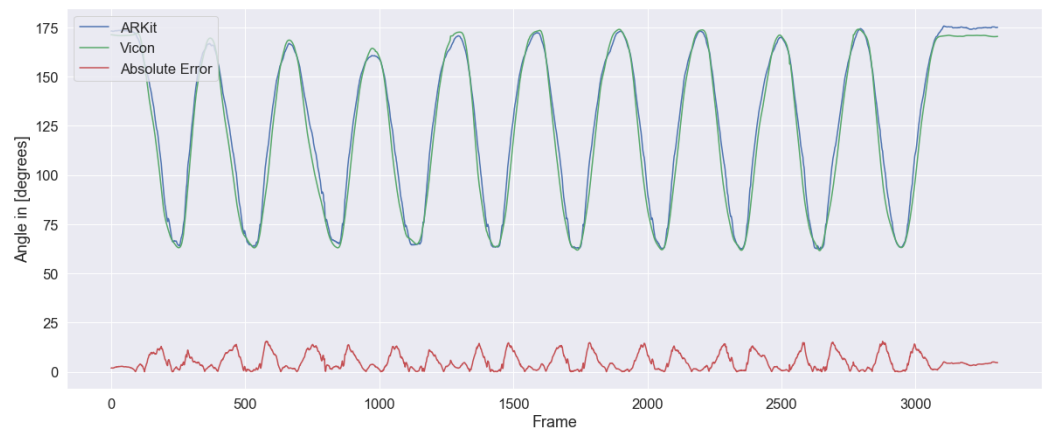


Figure 6. Left hip angle of one of the subjects in the Single Leg Deadlift exercise in degrees, which shows a nearly perfectly overlapping curves of the ARKit and Vicon data.

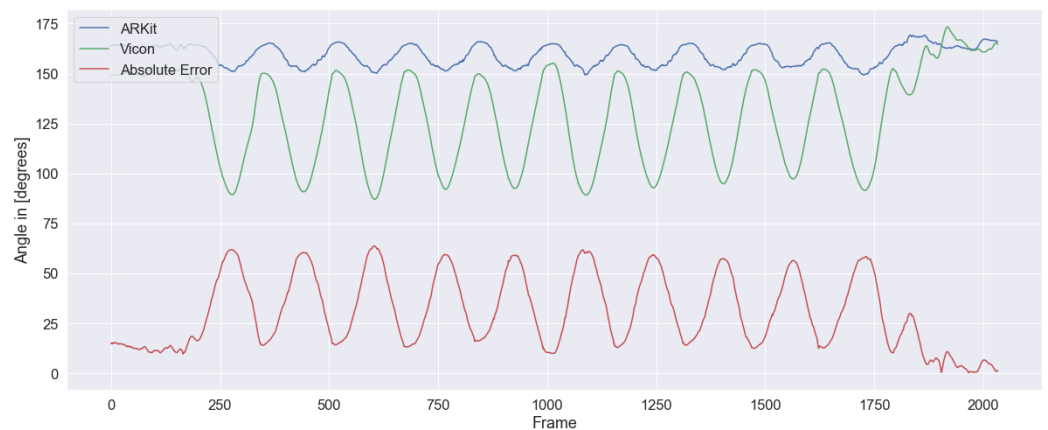


Figure 7. Left hip angle of one of the subjects in the Side Squat exercise in degrees. The plot shows that while the motion pattern is visible in both recordings, ARKit exposes a reduced amplitude and a shift on the y -axis.

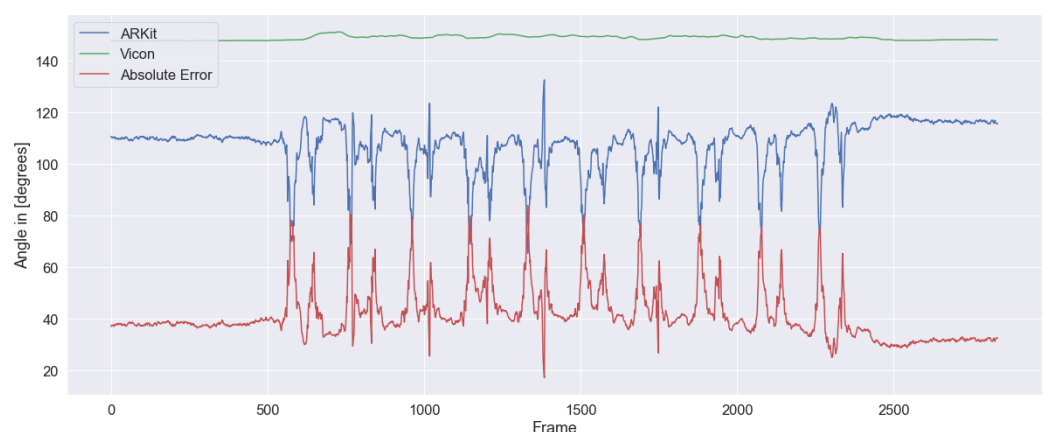


Figure 8. Right elbow angle of one of the subjects in the Squat exercise in degrees, which shows bad tracking quality with a lot of noise compared to the Vicon data.

Finding 1: ARKit is able to track the general progression of a movement with good accuracy but with significant deviations from the actual values measured by the Vicon system. The performance is influenced by external factors such as the performed motion.

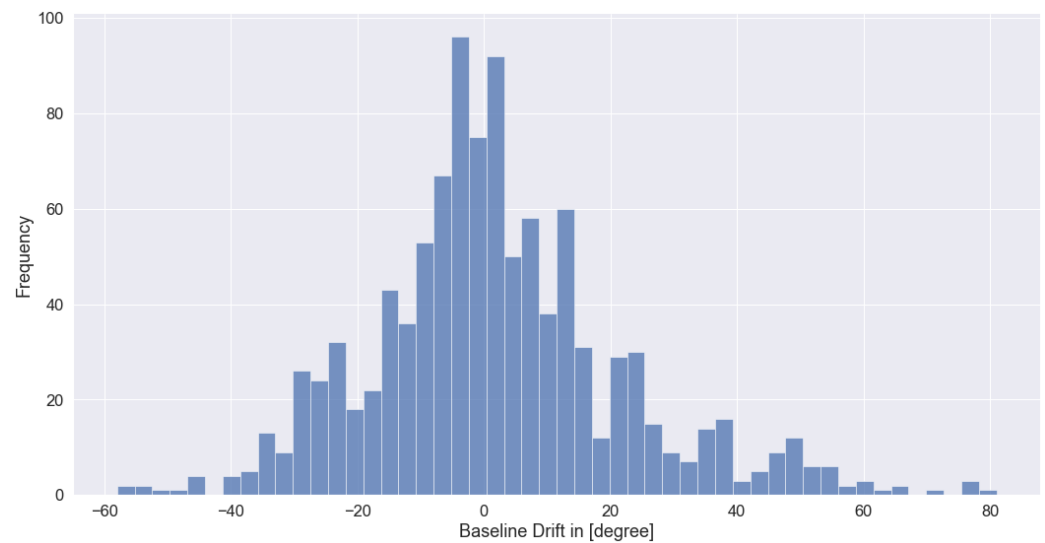


Figure 9. Results of the baseline drift analysis of the ARKit data. This is computed by minimizing the MAE by shifting the ARKit data vertically. The results show a normal distribution around 0, thus indicating no systematic baseline drift of the ARKit results.

4.2. RQ 2: Which Factors Influence ARKit's Motion Capture Results?

We performed factor analysis using Welch ANOVA, *t*-test analysis, and logistic regression on the dependent variable MAE to answer RQ2.

The MAE depended on the performed exercise. This dependency is visible when inspecting the respective boxplots of the MAE (Figure 10). Especially both Squat exercises (Squat, Side Squat) show significantly higher mean values than the other exercises. This observation is supported by the post-hoc analysis results of the ANOVA results. The logistic regression indicated an additional small influence of whether upper or lower body angles are considered. While the *t*-test showed an additional effect on whether the pelvic's center was moved during an exercise, this effect was not visible in the logistic regression. The impact of this factor remains inconclusive.

Finding 2: The factor analysis results show that the accuracy of ARKit's human motion capture mainly depends on the performed exercise.

While there is a slight difference between the frontal and side view data for both the wMAE and the SRCC, this difference is comparably small. The results of the side view show a 1.44° difference of the wMAE and a difference in the SRCC of 0.07, with the side view performing slightly better than the frontal view. These findings are supported by the factor analysis results, where no dependency of the view was measured. It also needs to be considered that the upper body angles in the side view only contained data of three subjects due to export problems, limiting the comparison's explanatory power.

Another aspect of the device's position influence is the visibility of specific body parts. Limited visibility of body joints, such as the left side of the body in the Front Lunge, Single Leg Deadlift, and Leg Extension Crunch, or the elbow joints in the Side Squat and Squat, is associated with a higher wMAE and worse correlation results, especially in the left elbow joint. Hidden joints often led to ARKit confusing the left and right body side for the respective joints, which caused unexpected peaks in the recorded data (Figure 11). The tracking of the upper body joints worked significantly better when other body parts did not hide them, as in the remaining three exercises Jumping Jacks, Lateral Arm Raise, and Reverse Fly.

Finding 3: When positioning the device, ensuring good visibility of the targeted joints improves the accuracy of the results.

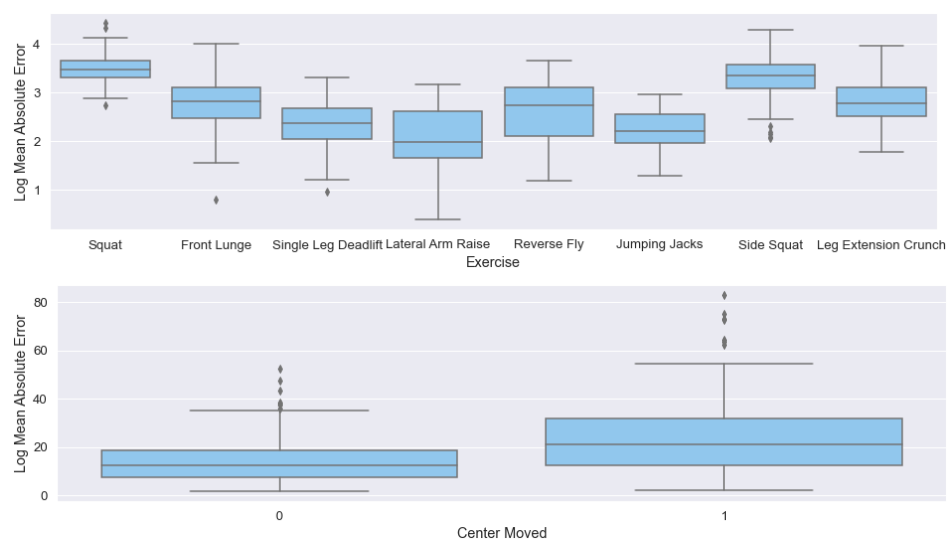


Figure 10. Boxplots representing the MAE in degrees on the logarithmic scale across all performed exercises and the *pelvic center moved* variable in the experiments. Both boxplots show significant differences in the mean and variance across the variables.

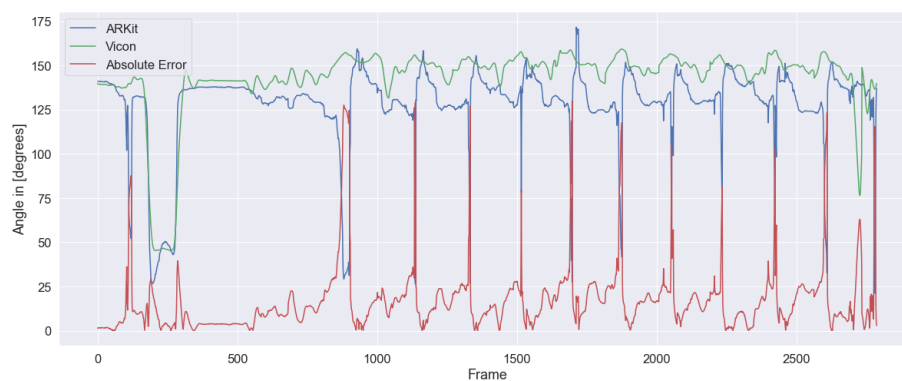


Figure 11. Left elbow angle of one of the subjects in the Single Leg Deadlift exercise, which shows several unexpected spikes during the execution. The spikes originate from ARKit incorrectly detecting the joint's position, most probably because of bad visibility of the elbow joint during the exercise.

5. Discussion

5.1. Factors Influencing ARKit's Performance

Based on the findings presented in Section 4, we identified several factors that influence the accuracy of ARKit's motion capture. The main requirement for good tracking is ensuring that the joints of interest are well visible to the camera and not hidden by other parts of the body during the movement. The exercise or motion itself is also of relevance. The results of the *t*-test hinted at a relevance of the coordinate system's stability during the exercise. However, this was not supported by the results of the logistic regression, so that the interpretation is unclear and requires further investigation.

The results of capturing human motion using ARKit could be influenced by several other factors, which were not further investigated within this research. This includes technical factors such as the device's processing power and additional sensors to improve the motion capture, the tracking environment such as lighting conditions or the background, or factors regarding the captured person, such as their clothing, body mass index, or skin color.

5.2. Bias of the Motion Capture Results

The upper body angles exposed a tendency of underestimation, and the results of the hips hinted at systematic overestimation as described in Section 3.1.2. Several values were

located close to -1 or 1 , which hints at a tendency to either systematic rather than cyclically occurring over- or underestimation. When aggregating the values for the different joints (Table 8), the results suggest that the upper body angles are underestimated, while the hip gets overestimated. The knee angles remain inconclusive with values relatively close to zero. They could hint at the mentioned cyclically occurring over- and underestimations or over- and underestimation based on the executed movement.

Table 8. The mean values of the ratio ME/MAE for the different joint angles.

Angle	Ratio ME/MAE
leftElbow	−0.46
rightElbow	−0.30
leftShoulder	−0.47
rightShoulder	−0.31
leftHip	0.59
rightHip	0.75
leftKnee	−0.19
rightKnee	0.01

5.3. Influence of the Tracked Joint Angle

The logistic regression results indicated a small, but significant effect of the lower body variable. These impressions are supported when inspecting the boxplot of the angles in the ME (Figure 12). The boxplot shows a tendency of underestimating the upper body angles, overestimating the hip angles, and a difference in the mean between the knee and hip angles. To investigate this effect, we performed the ANOVA analysis on the ME. We shifted the ME to only include positive values and applied the logarithmic transformation similar to our proceedings of the MAE as described in Section 2.6. The observed angles show an influence on the result ($\eta^2 = 0.26$, $p = 0.00$). Post-hoc analysis using Games-Howell supports the suggestions that the differences lie between the upper body angles and lower body angles and between the hip and knee angles (Appendix C, Table A2).

Interestingly, the exercise and movement of the hip center were the influencing factors for the MAE in contrast to the results of the ME. In the MAE, the difference between the angles is not observable anymore. The upper body error is mapped to a similar MAE as the lower body joints by only considering the absolute error (Figure 12). The ME for the whole dataset is -0.83° , meaning that overestimating the lower body joints and underestimating the upper body joints could be subject to error cancellation when considering the entire body. This effect could explain the MAE's dependency on the selected exercise while no dependency on the angle was observed.

The ANOVA results show an effect for the upper body variable and support the respective tendency of over- and underestimation. However, as explained in Section 2.6, the ME is prone to error cancellation effects. This unclear influence impacts the explanatory power, so we did not include these thoughts in the results and findings.

5.4. Impact of Incorrect Hip Detection

A commonly observed issue with the ARKit data were a reduced amplitude, and a baseline drift along the y -axis (see Figure 7), though the motion was tracked quite reliable. This issue was particularly the case for the lower body joints and led to a higher wMAE in those joints, but was also observed in other joints. In the screencasts of the recording, we often noticed that the detection of the hip joints was incorrect (Figure 13) and even varied during the execution of the exercise. Such shifts on the sagittal plane explain both the baseline drift and the amplitude reduction in the hip, knee, and shoulder angles, as all of them rely on the hip joints for their calculation. Especially from a side perspective, the hip joints allow for the most considerable deviations along the sagittal plane due to the amount of muscle and fat tissue around the pelvis. In the example of Figure 13, another issue aggravates the correct detection of the hip joints: the camera perspective was optimized

for tracking the legs' position, which in this case means that the right joint hides the left hip joint. This positioning implies that ARKit needs to rely on other body landmarks to estimate its position. Finding an optimal camera position in which all joints are completely visible might not be possible for all movements.

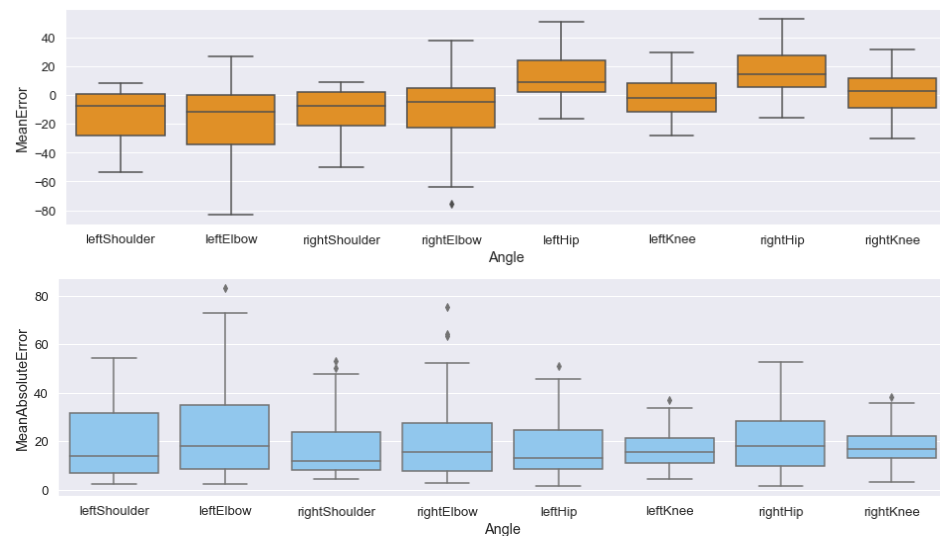


Figure 12. Boxplots representing the ME and MAE in degrees across all tracked angles in the experiments. The boxplots for the ME show a significant difference in the means of the upper and lower body angles, which is not visible for the MAE.

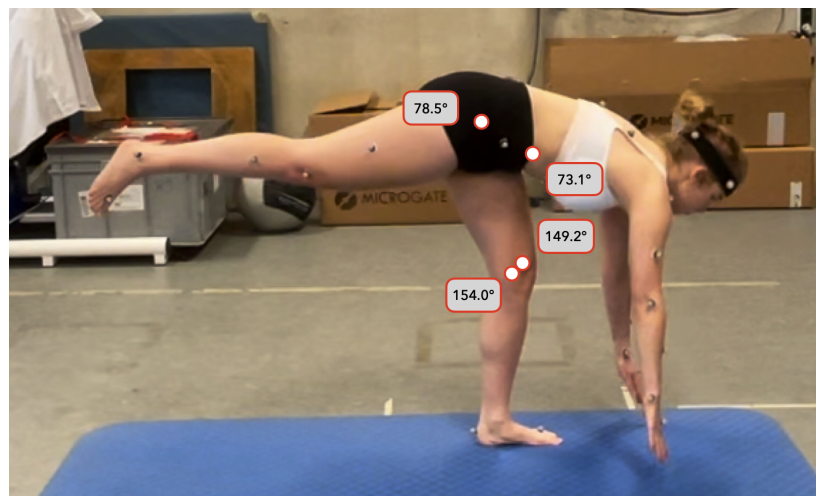


Figure 13. Exemplary screenshot of the frontal ARKit recording of one subject during the Single Leg Deadlift exercise, showing a bad detection of the hip joints and confusion of the knee joints.

5.5. Improving the ARKit Data during Post-Processing

The good correlation results opened up the question of whether it is possible to improve the ARKit motion capture data through post-processing to approximate the Vicon data. A systematic error concerning detecting the hip joints in a position too far anterior is a possible explanation and is subject to further investigation. If this is the case, both the baseline shift and the amplitude reduction could be corrected by applying a scale factor and shifting the data on the y -axis. Compensating the baseline shift would reduce the wMAE results by 7.61° and lead to more reliable and accurate results. However, no systematic error could be found when shifting the ARKit data along the y -axis by vertically shifting the ARKit data (Figure 9). The observed shift instead seems to be caused by other factors such as the incorrect detection of joints.

During the data analysis, we used a sliding window approach to maximizing the cross-correlation between the ARKit and Vicon data to compensate for possible time lags, as no synchronization of the iPads and the Vicon system was possible during the experiment. Possible reasons for lags are different hardware clocks and the delay of the body detection algorithm of the ARKit framework. The sliding window was set to a maximum of 120 frames, which equals approximately 2 s, only to allow reasonable shifts within the exercises and compensate for the lag caused by technical limitations. The approach was chosen to maximize the comparability between the results of the two systems. However, as the sliding window approach was applied individually to each angle, exercise, subject, and view, each configuration was shifted to its optimal result within the given time window. This approach does not consider potential lags within ARKit's motion capture, for example, a slower recognition of changes for some parts of the recognized body.

5.6. Comparing the Results of 2D and 3D Motion Capture Systems

As stated in the analysis of Sarafianos et al. [31], monocular video-based motion capture systems exhibit several limitations, which reduce their applicability to real-world scenarios. Among the most significant limitations are the ambiguities of the detected poses due to occlusion and distortion of the camera image caused by the camera's viewing angle and position [31], which is a relevant limitation in both 2D and 3D motion capture systems. In this research, we were able to show that ARKit, as an example for 3D motion capture systems supported by different smartphone sensors, is robust against a variation of 30° regarding the positioning of the device. The factor analysis did not expose an influence of the device position. However, poor visibility of joints still led to significant decreases in the accuracy of the measured angles. Mobile 3D motion capture frameworks based on monocular video data such as ARKit improve some of the limitations of 2D motion capture systems but cannot overcome them completely.

5.7. Potential Use Cases for Mobile 3D Motion Capture-Based Applications

The findings of this research raise the question of possible application areas for human motion capture using mobile 3D motion capture frameworks such as Apple ARKit. Referring to the three categories defined by Moeslund et al. [2], such frameworks could be applied to use cases in categories (2) interacting with software or (3) motion analysis for medical examinations or performance analysis, as it focuses on tracking single bodies rather than observing crowds. The results suggest that ARKit can track a motion's progression reliably but with relatively high error rates, depending on the joint of interest. Human motion capture using ARKit is further limited to a relatively small set of trackable joints. For example, the hand and toe joints are not actively tracked but calculated based on the angle and wrist joints, limiting the trackable joint angles to the shoulder, elbow, hip, and knee. However, mobile 3D motion capture frameworks are a promising technology for use cases that focus on tracking a specific motion of body parts rather than the exact joint position. Such use cases can be seen in category (2), such as interacting with software through gestures or other movements. Potential use cases in (3) include sports applications for amateurs or physiotherapy applications, which could focus on counting repetitions of a specific exercise. Depending on the motion and joint of interest, specific use cases relying on the exact joint position and angle data might be possible if the two main requirements for a good tracking presented at the beginning of this section can be met. For example, such use cases could include measuring the possible range of motion of a joint before and after a particular intervention and monitoring the progress in the medical field, or correcting the execution of a specific exercise in sports and physiotherapy applications. Using mobile 3D motion capture frameworks in these use cases would extend the usage of human motion capture technologies beyond professional settings and allow day-to-day usage at home, performed by consumers. ARKit and other mobile IPS systems enable new use cases, especially in mHealth, which were not possible with previous HMC systems. Our findings show how mobile 3D motion capture frameworks can be applied and how

mHealth applications could leverage the software for future applications. However, the limitations of 3D motion capture frameworks and ARKit's boundaries, in particular, need to be considered and should be evaluated before applying the technology to specific use cases.

5.8. Limitations

The design of this research includes several limitations. While the lab experiment produced a data set of over 1000 exercise executions, the data were collected from ten study participants only due to the restrictions caused by the ongoing COVID-19 pandemic. The limited number of participants might limit the *external validity* of this research. The participants' traits further limit the external validity. While covering heights between 156 cm and 198 cm, their body mass index was in a normal range. In addition, all participants had a lighter skin tone. The experiment was conducted in a laboratory with controlled background and lighting conditions.

Even though the study setup aimed at reducing possible influences on the study's *internal validity* which were not part of the observation, the impact of additional factors cannot be eliminated. Possible factors include the influence of the specific performance of the exercises by the subjects or the effect of the clothing worn. Furthermore, the subjects were recruited from the social surroundings of the researchers. They might not be representative of the whole population. The internal validity is further affected by the sliding window approach to compensate for the time lag due to missing clock synchronization and processing time. While the approach is limited to a maximum window of approximately two seconds, this shift could still have improved the results above the observable results. Additionally, the data set contained a reduced amount of exercise data for the upper body joints due to the export problems of the iPad on the side position. We applied the Welch ANOVA test to identify dependencies of the MAE instead of the ANOVA test, as the variance of the individual factors was not equally distributed. However, another prerequisite for (Welch) ANOVA and Welch *t*-test, normally distributed data, was only partially given for the MAE, even though the ANOVA analysis is said to be quite robust against this problem. We applied a logarithmic transformation to the data before performing the ANOVA and *t*-tests to overcome these limitations. Moreover, the observations used in (Welch) ANOVA should be independent of each other. In our experiment setup, the recording of angle motion happened simultaneously in all subjects and exercises. The observed angle deviations of the systems are expected to be independent. However, a poorly tracked angle might cause a higher risk to affect another angle's accuracy in a real-world scenario. Thus, the assumption of independent observations is hard to verify. Moreover, ARKit is only one example of a mobile 3D motion capture framework. Other frameworks rely on different technologies and algorithms and could exhibit different results and limitations.

6. Conclusions

This research evaluated mobile 3D motion capture based on the example of ARKit, Apple's framework for smartphone-based 3D motion capture. In contrast to existing monocular motion capture software, ARKit detects the human body in a 3-dimensional space instead of only two dimensions and augments its results by using smartphone sensor data such as IMU or depth data from the integrated LiDAR sensor. Our laboratory experiment, including ten participants, investigated ARKit's accuracy and influencing factors in eight body-weight exercises and compared it to the Vicon system, a gold standard for human motion capture. Our results provide evidence that mobile 3D motion capture frameworks can track the motion's progression with reasonable accuracy but with relatively high mean absolute error rates. The accuracy mainly depends on two factors: the visibility of the joints of interest and the observed motion. In contrast to 2D systems, the 3D motion capture framework exposed certain robustness against the positioning of the camera. However, similar limitations regarding the tracking of poorly visible joints remain.

Mobile 3D motion capture frameworks are promising and lightweight mobile technologies which could enable new use cases for human-computer interaction through motion

or application in health and medical fields. Their limitations, especially regarding the relatively high error rates compared to the gold standard system, need to be considered for each use case.

Author Contributions: Conceptualization, L.M.R., M.K., T.F. and S.M.J.; methodology, L.M.R., M.K. and S.M.J.; software, L.M.R.; validation, L.M.R., M.K., T.F. and S.M.J.; formal analysis, M.K. and L.M.R.; investigation, L.M.R. and T.F.; resources, L.M.R. and S.M.J.; data curation, M.K.; writing—original draft preparation, L.M.R., M.K. and T.F.; writing—review and editing, L.M.R., M.K. and S.M.J.; visualization, L.M.R. and M.K.; supervision, L.M.R. and S.M.J.; project administration, L.M.R.; funding acquisition, L.M.R. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by a grant from Software Campus through the German Federal Ministry of Education and Research, grant number 01IS17049.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Ethics Committee of the Technical University of Munich (Proposal 515/21 S on 19 August 2021). All participants were informed about the aims of the study and gave their consent about the publication of the anonymized data.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study. Written informed consent has been obtained from the subjects to publish this paper.

Data Availability Statement: All data is available on Zenodo [49].

Acknowledgments: We want to thank Florian Kreuzpointner for his support during the planning and execution of the study as well as the participants of the study.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

AOI	Angles of Interest
EMS	Electromagnetig Measurement Systems
FL	Front Lunge
HMC	Human Motion Capture
IPS	Image Processing Systems
IMU	Inertial Measurement Unit
JJ	Jumping Jacks
LAR	Lateral Arm Raise
LEC	Leg Extension Crunch
LE	Left Elbow
LH	Left Hip
LK	Left Knee
LS	Left Shoulder
MAE	Mean Absolute Error
ME	Mean Error
OMS	Optoelectronic Measurement Systems
PCC	Pearson Correlation Coefficient
RE	Right Elbow
RF	Reverse Fly
RH	Right Hip
RK	Right Knee
RS	Right Shoulder
S	Squat
SDK	Software Development Kit
SS	Side Squat

SLD Single Leg Deadlift
 ULS Ultrasonic Localization Systems
 wMAE Weighted Mean Absolute Error

Appendix A. Distributions of the Factors Used in the Welch ANOVA Analysis

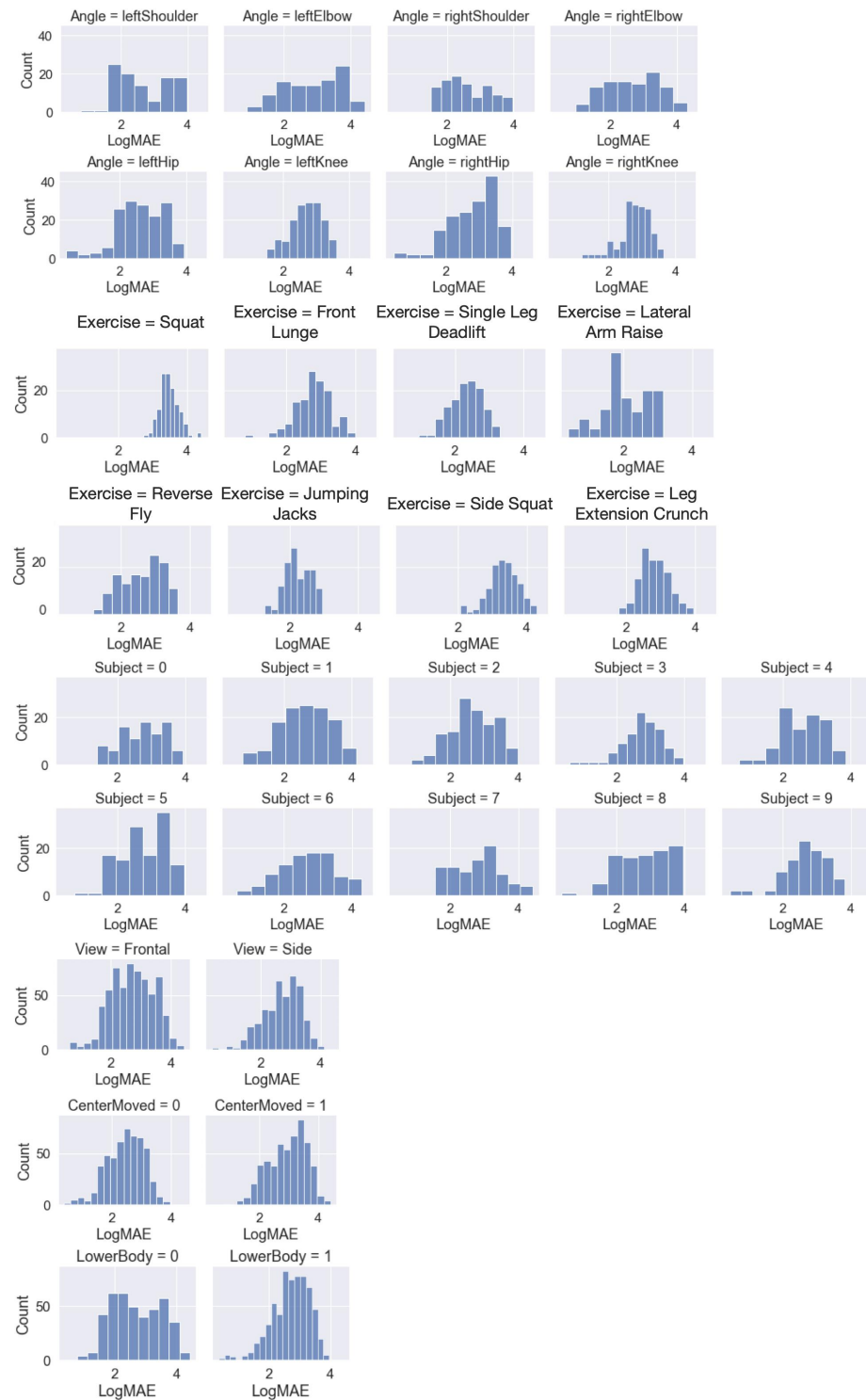


Figure A1. Distributions of the individual factors of the MAE on the logarithmic scale used in the factor analysis. Due to the transformation on the logarithmic scale, all factors are sufficiently close to a normal distribution, so that a factor analysis using Welch ANOVA/*t*-tests should be possible.

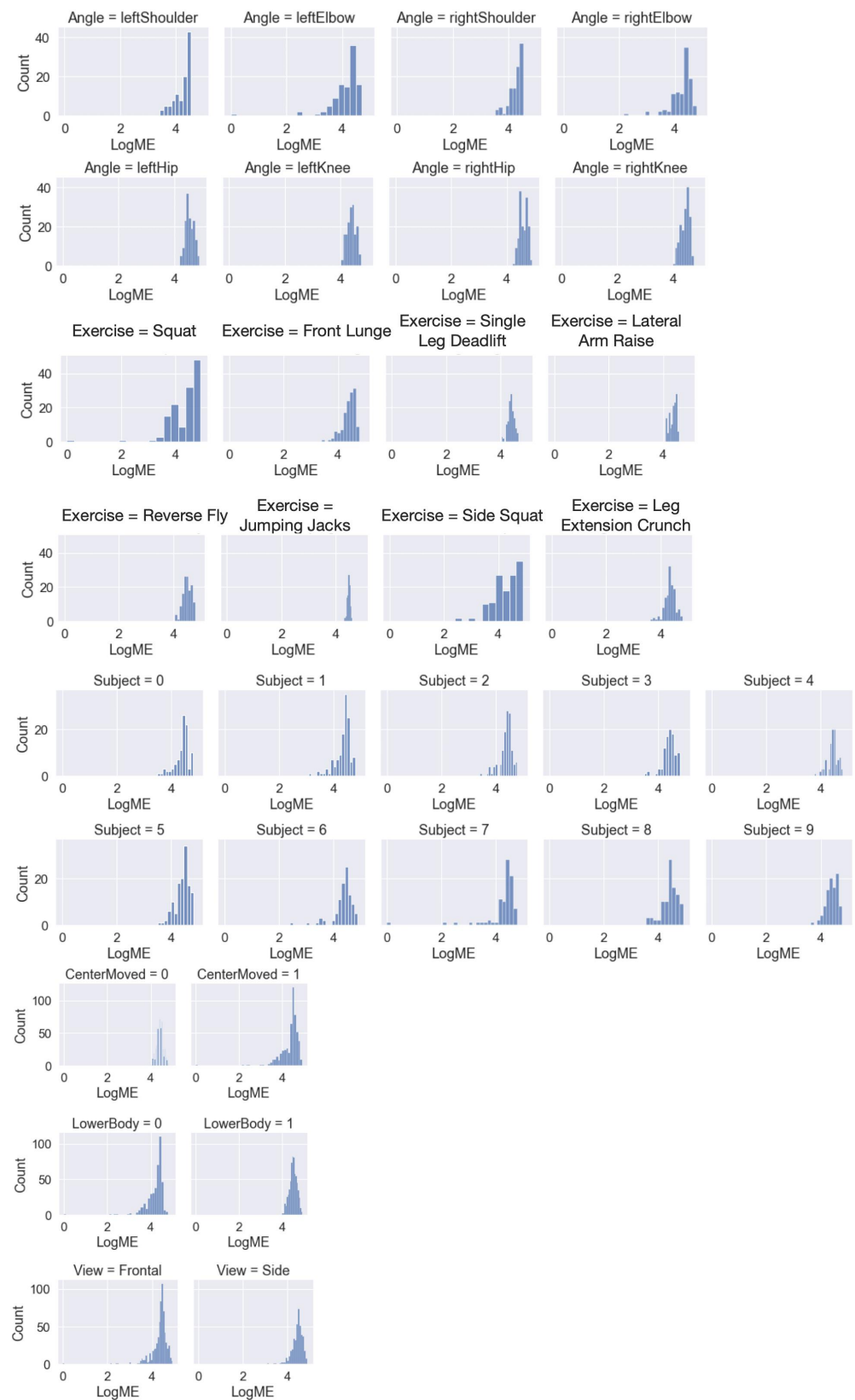


Figure A2. Distributions of the individual factors of the ME on the logarithmic scale used in the Welch ANOVA analysis. All of the factors show a distribution which is sufficiently close to a normal distribution so that an ANOVA analysis should be possible.

Appendix B. Bias

Exercise	Angle View	leftElbow	rightElbow	leftShoulder	rightShoulder	leftHip	rightHip	leftKnee	rightKnee
Front Lunge	Frontal	-27.98 ± 17.23	-20.72 ± 10.15	0.75 ± 3.59	-8.23 ± 5.2	-0.42 ± 8.77	19.84 ± 7.1	3.99 ± 8.63	9.04 ± 5.61
	Side	-28.08 ± 12.71	-30.91 ± 3.41	-10.38 ± 2.68	-14.8 ± 1.8	-3.52 ± 5.12	22.17 ± 3.25	10.16 ± 6.7	13.7 ± 4.29
Jumping Jacks	Frontal	2.58 ± 4.94	3.19 ± 3.71	2.35 ± 2.03	2.79 ± 2.28	6.29 ± 2.86	7.13 ± 2.6	-1.08 ± 3.85	0.45 ± 4.0
	Side	8.39 ± 1.52	6.48 ± 1.49	0.5 ± 2.51	5.6 ± 0.76	7.17 ± 3.32	9.58 ± 2.84	1.79 ± 4.62	2.77 ± 4.27
Lateral Arm Raise	Frontal	-5.46 ± 5.62	-6.03 ± 3.76	1.26 ± 3.42	3.04 ± 3.72	3.9 ± 4.11	3.88 ± 3.73	-17.53 ± 4.21	-17.57 ± 3.29
	Side	-2.48 ± 1.92	-3.48 ± 2.27	-0.88 ± 2.38	3.24 ± 1.98	1.59 ± 4.95	6.62 ± 3.62	-17.69 ± 4.48	-17.33 ± 3.83
Leg Extension Crunch	Frontal	-0.22 ± 11.99	16.22 ± 7.8	-28.58 ± 9.15	-14.62 ± 5.89	-3.52 ± 3.39	4.17 ± 4.25	-9.32 ± 4.61	-6.28 ± 5.0
	Side	9.42 ± 16.08	24.91 ± 12.09	-35.27 ± 3.82	-19.19 ± 6.97	-4.8 ± 5.41	8.31 ± 5.88	-13.31 ± 4.26	-15.84 ± 7.75
Reverse Fly	Frontal	-2.83 ± 7.09	3.87 ± 5.78	-5.45 ± 6.48	-3.01 ± 5.7	24.94 ± 6.35	27.43 ± 6.75	5.26 ± 12.58	5.78 ± 12.27
	Side	-0.65 ± 5.28	4.61 ± 2.58	-11.51 ± 2.06	-2.23 ± 4.3	22.11 ± 7.19	20.94 ± 7.54	6.77 ± 12.56	3.64 ± 10.85
Side Squat	Frontal	-46.02 ± 14.64	-41.15 ± 15.21	-38.13 ± 11.12	-33.23 ± 9.59	22.55 ± 7.62	36.39 ± 5.22	-12.39 ± 9.89	15.85 ± 6.24
	Side	-46.08 ± 5.2	-27.6 ± 4.03	-28.62 ± 3.56	-20.89 ± 1.9	26.08 ± 8.27	30.22 ± 3.97	-11.36 ± 10.51	11.55 ± 4.1
Single Leg Deadlift	Frontal	-18.27 ± 9.06	-3.3 ± 8.26	-3.24 ± 5.66	-2.05 ± 5.62	9.28 ± 3.17	-3.04 ± 6.62	-3.36 ± 3.67	-4.29 ± 7.56
	Side	-12.78 ± 0.73	-4.96 ± 1.74	-1.72 ± 5.84	2.67 ± 4.01	11.4 ± 5.31	-4.16 ± 7.3	-4.21 ± 4.01	-5.72 ± 7.36
Squat	Frontal	-43.14 ± 18.84	-36.95 ± 17.32	-34.59 ± 11.14	-32.9 ± 10.54	34.96 ± 6.65	37.05 ± 6.75	14.72 ± 6.47	15.85 ± 7.71
	Side	-44.0 ± 11.41	-31.72 ± 9.74	-30.3 ± 6.29	-23.41 ± 2.55	31.45 ± 5.14	29.88 ± 4.98	14.16 ± 5.0	9.89 ± 6.24

Figure A3. Pivot Table of the average Mean Error (ME) distributed over the eight exercises and the eight tracked angles, each measured from the two iPad perspectives *Frontal* and *Side*. The dashed boxes indicate which joints were specifically targeted by the respective exercise. The heatmap visualizes the performance of the individual joints per exercise, with darker purple color hinting at underestimation and darker orange color hinting at overestimation. Values closer to zero either indicate good performance or error cancellation.

Exercise	Angle View	leftElbow	rightElbow	leftShoulder	rightShoulder	leftHip	rightHip	leftKnee	rightKnee
Front Lunge	Frontal	-0.79	-0.87	0.09	-0.67	-0.04	0.98	0.29	0.65
	Side	-0.77	-0.92	-0.71	-0.84	-0.28	1.00	0.58	0.76
Jumping Jacks	Frontal	0.34	0.45	0.36	0.37	0.75	0.86	-0.07	0.03
	Side	0.92	0.70	0.08	0.72	0.76	0.96	0.12	0.20
Lateral Arm Raise	Frontal	-0.73	-0.82	0.19	0.45	0.78	0.81	-0.99	-1.00
	Side	-0.66	-0.64	-0.16	0.57	0.32	0.97	-1.00	-0.99
Leg Extension Crunch	Frontal	-0.01	0.83	-0.89	-0.84	-0.32	0.36	-0.63	-0.38
	Side	0.48	0.96	-0.97	-0.90	-0.34	0.51	-0.78	-0.79
Reverse Fly	Frontal	-0.32	0.38	-0.46	-0.28	0.99	0.99	0.34	0.39
	Side	-0.08	0.31	-0.78	-0.26	0.99	1.00	0.44	0.30
Side Squat	Frontal	-0.98	-0.97	-0.91	-0.92	1.00	1.00	-0.75	0.60
	Side	-0.98	-0.99	-0.95	-0.88	1.00	1.00	-0.70	0.54
Single Leg Deadlift	Frontal	-0.87	-0.38	-0.28	-0.19	0.88	-0.21	-0.39	-0.46
	Side	-0.96	-0.81	-0.18	0.37	0.93	-0.28	-0.52	-0.57
Squat	Frontal	-0.97	-0.98	-0.88	-0.88	0.99	0.99	0.49	0.52
	Side	-0.97	-1.00	-0.83	-0.84	0.98	0.99	0.51	0.42

Figure A4. Pivot Table of the ratio of the ME divided by the MAE distributed over the eight exercises and the eight tracked angles, each measured from the two iPad perspectives *Frontal* and *Side*. The dashed boxes indicate which joints were specifically targeted by the respective exercise. The heatmap visualizes the performance of the individual joints per exercise. Values close to zero indicate either good performance of the tracking or over- and underestimation canceling each other out. Values closer to -1 and 1 hint at systematic under- and overestimation in the specific configuration.

Appendix C. ANOVA Post-Hoc Analysis

Appendix C.1. Mean Absolute Error

Table A1. The results of the ANOVA Post-hoc analysis of the MAE for the eight exercises Front Lunge (FL), Jumping Jacks (JJ), Lateral Arm Raise (LAR), Leg Extension Crunch (LEC), Reverse Fly (RF), Side Squat (SS), Single Leg Deadlift (SLD), and Squat (S).

A	B	Mean(A)	Mean(B)	Diff	se	T	df	p	η^2
FL	JJ	2.78	2.25	0.53	0.06	9.69	242.95	0.00	0.26
FL	LAR	2.78	2.04	0.74	0.07	10.05	240.85	0.00	0.28
FL	LEC	2.78	2.81	-0.03	0.06	-0.49	254.87	1.00	0.00
FL	RF	2.78	2.61	0.17	0.07	2.53	254.89	0.19	0.02
FL	SS	2.78	3.32	-0.54	0.06	-9.35	257.60	0.00	0.25
FL	SLD	2.78	2.33	0.45	0.06	7.51	253.76	0.00	0.18
FL	S	2.78	3.49	-0.71	0.05	-14.17	204.84	0.00	0.43
JJ	LAR	2.25	2.04	0.21	0.07	3.12	204.18	0.04	0.04
JJ	LEC	2.25	2.81	-0.56	0.05	-11.29	258.38	0.00	0.33
JJ	RF	2.25	2.61	-0.36	0.06	-5.85	221.58	0.00	0.12
JJ	SS	2.25	3.32	-1.07	0.05	-21.28	255.86	0.00	0.63
JJ	SLD	2.25	2.33	-0.08	0.05	-1.51	238.68	0.80	0.01
JJ	S	2.25	3.49	-1.24	0.04	-30.34	241.51	0.00	0.78
LAR	LEC	2.04	2.81	-0.77	0.07	-11.00	219.24	0.00	0.31
LAR	RF	2.04	2.61	-0.57	0.08	-7.24	236.94	0.00	0.17
LAR	SS	2.04	3.32	-1.28	0.07	-18.19	224.12	0.00	0.56
LAR	SLD	2.04	2.33	-0.29	0.07	-4.03	230.32	0.00	0.06
LAR	S	2.04	3.49	-1.45	0.06	-22.61	173.71	0.00	0.66
LEC	RF	2.81	2.61	0.20	0.06	3.13	236.94	0.04	0.04
LEC	SS	2.81	3.32	-0.52	0.05	-9.96	261.64	0.00	0.26
LEC	SLD	2.81	2.33	0.48	0.06	8.66	249.27	0.00	0.23
LEC	S	2.81	3.49	-0.68	0.04	-15.40	226.49	0.00	0.47
RF	SS	2.61	3.32	-0.71	0.06	-11.10	241.49	0.00	0.32
RF	SLD	2.61	2.33	0.28	0.07	4.22	244.66	0.00	0.07
RF	S	2.61	3.49	-0.88	0.06	-15.39	186.05	0.00	0.47
SS	SLD	3.32	2.33	0.99	0.06	17.69	251.45	0.00	0.55
SS	S	3.32	3.49	-0.17	0.04	-3.63	221.60	0.01	0.05
SLD	S	2.33	3.49	-1.16	0.05	-24.28	200.97	0.00	0.70

Appendix C.2. Mean Error

Table A2. The results of the ANOVA Post-hoc analysis of the ME for the eight angles left elbow (LE), left hip (LH), left knee (LK), left shoulder (LS), right elbow (RE), right hip (RH), right knee (RK), and right shoulder (RS).

A	B	Mean(A)	Mean(B)	Diff	se	T	df	p	η^2
LE	LH	4.10	4.55	-0.45	0.06	-7.73	110.78	0.00	0.19
LE	LK	4.10	4.40	-0.29	0.06	-5.03	111.83	0.00	0.09
LE	LS	4.10	4.22	-0.11	0.06	-1.78	148.18	0.63	0.01
LE	RE	4.10	4.25	-0.15	0.07	-2.14	177.59	0.39	0.02
LE	RH	4.10	4.60	-0.50	0.06	-8.54	110.20	0.00	0.23
LE	RK	4.10	4.44	-0.34	0.06	-5.74	111.85	0.00	0.12
LE	RS	4.10	4.27	-0.17	0.06	-2.69	134.33	0.14	0.03
LH	LK	4.55	4.40	0.16	0.02	9.12	315.23	0.00	0.21
LH	LS	4.55	4.22	0.34	0.03	11.13	138.85	0.00	0.33
LH	RE	4.55	4.25	0.30	0.04	7.63	121.90	0.00	0.19
LH	RH	4.55	4.60	-0.05	0.02	-2.81	313.63	0.10	0.02
LH	RK	4.55	4.44	0.12	0.02	6.71	315.19	0.00	0.12
LH	RS	4.55	4.27	0.28	0.03	11.03	155.72	0.00	0.33
LK	LS	4.40	4.22	0.18	0.03	5.92	143.19	0.00	0.12
LK	RE	4.40	4.25	0.15	0.04	3.68	124.27	0.01	0.05

Table A2. Cont.

A	B	Mean(A)	Mean(B)	Diff	se	T	df	p	η^2
LK	RH	4.40	4.60	−0.20	0.02	−11.99	313.81	0.00	0.31
LK	RK	4.40	4.44	−0.04	0.02	−2.34	318.00	0.28	0.02
LK	RS	4.40	4.27	0.13	0.03	4.91	161.84	0.00	0.09
LS	RE	4.22	4.25	−0.03	0.05	−0.72	187.27	1.00	0.00
LS	RH	4.22	4.60	−0.38	0.03	−12.72	136.43	0.00	0.39
LS	RK	4.22	4.44	−0.22	0.03	−7.26	143.28	0.00	0.17
LS	RS	4.22	4.27	−0.05	0.04	−1.45	196.84	0.83	0.01
RE	RH	4.25	4.60	−0.35	0.04	−8.82	120.58	0.00	0.24
RE	RK	4.25	4.44	−0.19	0.04	−4.71	124.32	0.00	0.08
RE	RS	4.25	4.27	−0.02	0.04	−0.41	167.96	1.00	0.00
RH	RK	4.60	4.44	0.16	0.02	9.54	313.75	0.00	0.22
RH	RS	4.60	4.27	0.33	0.03	12.90	152.29	0.00	0.40
RK	RS	4.44	4.27	0.17	0.03	6.49	161.97	0.00	0.14

References

- Moeslund, T.B.; Granum, E. A Survey of Computer Vision-Based Human Motion Capture. *Comput. Vis. Image Underst.* **2001**, *81*, 231–268. [\[CrossRef\]](#)
- Moeslund, T.B.; Hilton, A.; Krüger, V. A survey of advances in vision-based human motion capture and analysis. *Comput. Vis. Image Underst.* **2006**, *104*, 90–126. [\[CrossRef\]](#)
- Chiari, L.; Croce, U.D.; Leardini, A.; Cappozzo, A. Human movement analysis using stereophotogrammetry. *Gait Posture* **2005**, *21*, 197–211. [\[CrossRef\]](#) [\[PubMed\]](#)
- Elliott, B.; Alderson, J. Laboratory versus field testing in cricket bowling: A review of current and past practice in modelling techniques. *Sports Biomech.* **2007**, *6*, 99–108. [\[CrossRef\]](#) [\[PubMed\]](#)
- Carse, B.; Meadows, B.; Bowers, R.; Rowe, P. Affordable clinical gait analysis: An assessment of the marker tracking accuracy of a new low-cost optical 3D motion analysis system. *Physiotherapy* **2013**, *99*, 347–351. [\[CrossRef\]](#) [\[PubMed\]](#)
- McLean, S.G. Evaluation of a two dimensional analysis method as a screening and evaluation tool for anterior cruciate ligament injury. *Br. J. Sports Med.* **2005**, *39*, 355–362. [\[CrossRef\]](#)
- van der Kruk, E.; Reijne, M.M. Accuracy of human motion capture systems for sport applications; state-of-the-art review. *Eur. J. Sport Sci.* **2018**, *18*, 806–819. [\[CrossRef\]](#)
- Belyea, B.C.; Lewis, E.; Gabor, Z.; Jackson, J.; King, D.L. Validity and Intrarater Reliability of a 2-Dimensional Motion Analysis Using a Handheld Tablet Compared With Traditional 3-Dimensional Motion Analysis. *J. Sport Rehabil.* **2015**, *24*, 1014–1019. [\[CrossRef\]](#)
- Paul, S.S.; Lester, M.E.; Foreman, K.B.; Dibble, L.E. Validity and Reliability of Two-Dimensional Motion Analysis for Quantifying Postural Deficits in Adults With and Without Neurological Impairment. *Anat. Rec.* **2016**, *299*, 1165–1173. [\[CrossRef\]](#)
- Springer, S.; Seligmann, G.Y. Validity of the Kinect for Gait Assessment: A Focused Review. *Sensors* **2016**, *16*, 194. [\[CrossRef\]](#)
- Puh, U.; Hoehlein, B.; Deutsch, J.E. Validity and Reliability of the Kinect for Assessment of Standardized Transitional Movements and Balance. *Phys. Med. Rehabil. Clin. N. Am.* **2019**, *30*, 399–422. [\[CrossRef\]](#)
- Schärer, C.; Siebenthal, L.V.; Lomax, I.; Gross, M.; Taube, W.; Hübner, K. Simple Assessment of Height and Length of Flight in Complex Gymnastic Skills: Validity and Reliability of a Two-Dimensional Video Analysis Method. *Appl. Sci.* **2019**, *9*, 3975. [\[CrossRef\]](#)
- Alahmari, A.; Herrington, L.; Jones, R. Concurrent validity of two-dimensional video analysis of lower-extremity frontal plane of movement during multidirectional single-leg landing. *Phys. Ther. Sport* **2020**, *42*, 40–45. [\[CrossRef\]](#)
- Vicon Motion Capture Systems. Available online: <https://www.vicon.com> (accessed on 26 January 2022).
- Qualisys Motion Capture Systems. Available online: <https://www.qualisys.com> (accessed on 26 January 2022).
- Xsens Motion Capture Systems. Available online: <https://www.xsens.com> (accessed on 26 January 2022).
- Perception Neuron Motion Capture. Available online: <https://neuronmocap.com/> (accessed on 26 January 2022).
- Stelzer, A.; Pourvoyeur, K.; Fischer, A. Concept and Application of LPM—A Novel 3-D Local Position Measurement System. *IEEE Trans. Microw. Theory Tech.* **2004**, *52*, 2664–2669. [\[CrossRef\]](#)
- OpenPose: Real-Time Multi-Person Keypoint Detection Library for Body, Face, Hands, and Foot Estimation. Available online: <https://github.com/CMU-Perceptual-Computing-Lab/openpose> (accessed on 26 January 2022).
- ARKit: Capturing Body Motion in 3D. Available online: https://developer.apple.com/documentation/arkit/content_anchors/capturing_body_motion_in_3d (accessed on 26 January 2022).
- Vision: Detecting Human Body Poses in Images. Available online: https://developer.apple.com/documentation/vision/detecting_human_body_poses_in_images (accessed on 26 January 2022).
- TensorFlow Pose Estimate. Available online: https://www.tensorflow.org/lite/examples/pose_estimation/overview (accessed on 26 January 2022).

23. Cao, Z.; Hidalgo, G.; Simon, T.; Wei, S.E.; Sheikh, Y. OpenPose: Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *43*, 172–186. [CrossRef] [PubMed]
24. Whittle, M.W. Clinical gait analysis: A review. *Hum. Mov. Sci.* **1996**, *15*, 369–387. [CrossRef]
25. Oyeboode, O.; Ndulue, C.; Alhasani, M.; Orji, R. Persuasive Mobile Apps for Health and Wellness: A Comparative Systematic Review. In *Lecture Notes in Computer Science*; Springer International Publishing: Zurich, Switzerland, 2020; pp. 163–181. [CrossRef]
26. research2guidance. Number of Downloads of mHealth Apps Worldwide from 2013 to 2018 (in Billions) [Graph]. 2018. Available online: <https://de-statista-com/statistik/daten/studie/695434/umfrage/nummer-der-weltweiten-downloads-von-mhealth-apps/> (accessed on 26 January 2022).
27. Schoeppe, S.; Alley, S.; Lippevelde, W.V.; Bray, N.A.; Williams, S.L.; Duncan, M.J.; Vandelanotte, C. Efficacy of interventions that use apps to improve diet, physical activity and sedentary behaviour: A systematic review. *Int. J. Behav. Nutr. Phys. Act.* **2016**, *13*, 127. [CrossRef]
28. Boulos, M.N.K.; Brewer, A.C.; Karimkhani, C.; Buller, D.B.; Dellavalle, R.P. Mobile medical and health apps: State of the art, concerns, regulatory control and certification. *Online J. Public Health Inform.* **2014**, *5*, 229. [CrossRef]
29. Lopes, T.J.A.; Ferrari, D.; Ioannidis, J.; Simic, M.; Azevedo, F.M.D.; Pappas, E. Reliability and Validity of Frontal Plane Kinematics of the Trunk and Lower Extremity Measured with 2-Dimensional Cameras During Athletic Tasks: A Systematic Review with Meta-analysis. *J. Orthop. Sports Phys. Ther.* **2018**, *48*, 812–822. [CrossRef]
30. Zago, M.; Luzzago, M.; Marangoni, T.; Cecco, M.D.; Tarabini, M.; Galli, M. 3D Tracking of Human Motion Using Visual Skeletonization and Stereoscopic Vision. *Front. Bioeng. Biotechnol.* **2020**, *8*, 181. [CrossRef]
31. Sarafianos, N.; Boteanu, B.; Ionescu, B.; Kakadiaris, I.A. 3D Human pose estimation: A review of the literature and analysis of covariates. *Comput. Vis. Image Underst.* **2016**, *152*, 1–20. [CrossRef]
32. Cao, Z.; Simon, T.; Wei, S.E.; Sheikh, Y. *Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields*; CVPR: Prague, Czech Republic, 2017.
33. Simon, T.; Joo, H.; Matthews, I.; Sheikh, Y. *Hand Keypoint Detection in Single Images Using Multiview Bootstrapping*; CVPR: Prague, Czech Republic, 2017.
34. Wei, S.E.; Ramakrishna, V.; Kanade, T.; Sheikh, Y. *Convolutional Pose Machines*; CVPR: Prague, Czech Republic, 2016. [CrossRef]
35. D’Antonio, E.; Taborri, J.; Palermo, E.; Rossi, S.; Patane, F. A markerless system for gait analysis based on OpenPose library. In Proceedings of the 2020 IEEE International Instrumentation and Measurement Technology Conference (I2MTC), Dubrovnik, Croatia, 25–28 May 2020; IEEE: Dubrovnik, Croatia, 2020. [CrossRef]
36. Ota, M.; Tateuchi, H.; Hashiguchi, T.; Kato, T.; Ogino, Y.; Yamagata, M.; Ichihashi, N. Verification of reliability and validity of motion analysis systems during bilateral squat using human pose tracking algorithm. *Gait Posture* **2020**, *80*, 62–67. [CrossRef] [PubMed]
37. Nakano, N.; Sakura, T.; Ueda, K.; Omura, L.; Kimura, A.; Iino, Y.; Fukashiro, S.; Yoshioka, S. Evaluation of 3D Markerless Motion Capture Accuracy Using OpenPose With Multiple Video Cameras. *Front. Sports Act. Living* **2020**, *2*, 50. [CrossRef] [PubMed]
38. MediaPipe Pose. Available online: <https://google.github.io/mediapipe/solutions/pose.html> (accessed on 18 March 2022).
39. Bazarevsky, V.; Grishchenko, I.; Raveendran, K.; Zhu, T.; Zhang, F.; Grundmann, M. BlazePose: On-device Real-time Body Pose tracking. *arXiv* **2020**, arXiv:2006.10204.
40. Zhou, X.; Leonardos, S.; Hu, X.; Daniilidis, K. 3D shape estimation from 2D landmarks: A convex relaxation approach. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–15 June 2015; IEEE: Boston, MA, USA, 2015; pp. 4447–4455. [CrossRef]
41. Zhou, X.; Zhu, M.; Leonardos, S.; Derpanis, K.G.; Daniilidis, K. Sparseness Meets Deepness: 3D Human Pose Estimation from Monocular Video. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; IEEE: Las Vegas, NV, USA, 2016; pp. 4966–4975. [CrossRef]
42. Akhter, I.; Black, M.J. Pose-conditioned joint angle limits for 3D human pose reconstruction. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–15 June 2015; IEEE: Boston, MA, USA, 2015; pp. 1446–1455. [CrossRef]
43. Ma, F.; Cavalheiro, G.V.; Karaman, S. Self-Supervised Sparse-to-Dense: Self-Supervised Depth Completion from LiDAR and Monocular Camera. In Proceedings of the 2019 International Conference on Robotics and Automation (ICRA), Montreal, QC, Canada, 20–24 May 2019; IEEE: Montreal, QC, Canada, 2019; pp. 3288–3295. [CrossRef]
44. Reimer, L.M.; Weigel, S.; Ehrenstorfer, F.; Adikari, M.; Birkle, W.; Jonas, S. Mobile Motion Tracking for Disease Prevention and Rehabilitation Using Apple ARKit. In *Studies in Health Technology and Informatics*; Hayn, D., Schreier, G., Baumgartner, M., Eds.; IOS Press: Amsterdam, The Netherlands, 2021. [CrossRef]
45. Basiratizadeh, S.; Lemaire, E.D.; Baddour, N. Augmented Reality Approach for Marker-based Posture Measurement on Smartphones. In Proceedings of the 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Montreal, QC, Canada, 20–24 July 2020; IEEE: Montreal, QC, Canada, 2020; pp. 4612–4615. [CrossRef]
46. Full Body Modeling with Plug-In Gait. Available online: <https://docs.vicon.com/display/Nexus212/Full+body+modeling+with+Plug-in+Gait> (accessed on 26 January 2022).

47. Schmider, E.; Ziegler, M.; Danay, E.; Beyer, L.; Bühner, M. Is It Really Robust?: Reinvestigating the Robustness of ANOVA Against Violations of the Normal Distribution Assumption. *Methodology* **2010**, *6*, 147–151. [[CrossRef](#)]
48. Games, P.A.; Howell, J.F. Pairwise multiple comparison procedures with unequal n's and/or variances: A Monte Carlo study. *J. Educ. Stat.* **1976**, *1*, 113–125.
49. Reimer, L.M.; Kapsecker, M.; Fukushima, T.; Jonas, S.M. A Dataset for Evaluating 3D Motion Captured Synchronously by ARKit and Vicon. *ZENODO* **2022**. [[CrossRef](#)]