Titre: Title:	Evaluation of Sit-to-Stand Exercises in Rehabilitation Using a Cost-Effective Markerless Motion Capture
Auteur: Author:	Louis-Simon Guay
Date:	2025
Type:	Mémoire ou thèse / Dissertation or Thesis
Référence: Citation:	Guay, L.-S. (2025). Evaluation of Sit-to-Stand Exercises in Rehabilitation Using a Cost-Effective Markerless Motion Capture [Mémoire de maîtrise, Polytechnique Montréal]. PolyPublie. https://publications.polymtl.ca/67711/

URL de PolyPublie: PolyPublie URL:	https://publications.polymtl.ca/67711/
Directeurs de recherche: Advisors:	Abolfazl Mohebbi, & Sylvie Nadeau
Programme: Program:	Génie biomédical

POLYTECHNIQUE MONTRÉAL

affiliée à l'Université de Montréal

Evaluation of Sit-to-Stand Exercises in Rehabilitation Using Cost-Effective Markerless Motion Capture

LOUIS-SIMON GUAY

Institut de génie biomédical

Mémoire présenté en vue de l'obtention du diplôme de *Maîtrise ès sciences appliquées*

Génie biomédical

Juillet 2025

POLYTECHNIQUE MONTRÉAL

affiliée à l'Université de Montréal

Ce mémoire intitulé :

Evaluation of Sit-to-Stand Exercises in Rehabilitation Using Cost-Effective Markerless Motion Capture

présenté par **Louis-Simon GUAY**

en vue de l'obtention du diplôme de *Maîtrise ès sciences appliquées*

a été dûment accepté par le jury d'examen constitué de :

René MAYER, président
Abolfazl MOHEBBI, membre et directeur de recherche
Sylvie NADEAU, membre et codirectrice de recherche
Lama SÉOUD, membre

ACKNOWLEDGEMENTS

I would like to thank my supervisors, Abolfazl Mohebbi and Sylvie Nadeau, for their guidance, time, and support throughout this project. I am also grateful to Youssef El Khamlichi for his assistance with the Vicon experiment and data processing, and to Lama Séoud for providing the 3D cameras, along with valuable insight related to human pose estimation.

RÉSUMÉ

Ce mémoire évalue le potentiel clinique de systèmes de capture de mouvement sans marqueurs à faible coût pour l'évaluation du mouvement assis-debout (StS) en réadaptation. Deux caméras 3D grand public, l'Intel RealSense D455 (stéréo) et la L515 (temps de vol), ont été comparées à un système de référence Vicon. Une architecture logicielle personnalisée utilisant le logiciel Nuitrack SDK a été développée pour extraire les angles articulaires et les vitesses à partir de vidéos RGB-D. L'expérimentation a été réalisée avec un seul participant effectuant des répétitions StS, captées simultanément par les deux caméras 3D et par le système Vicon. La précision du système a été évaluée en répétant l'analyse 30 fois sur chaque enregistrement, afin de mesurer la stabilité des résultats obtenus par estimation de pose. L'exactitude a ensuite été déterminée en comparant les mesures angulaires aux données de référence du Vicon pour un ensemble de 11 conditions expérimentales couvrant des variations réalistes de vitesse d'exécution, de posture, de symétrie et de type de vêtement. Les résultats montrent que la caméra L515 produit des trajectoires cohérentes pour plusieurs angles articulaires dans des scénarios StS standards, avec un coefficient de concordance (CCC) moyen de 0.91. Toutefois, d'autres indicateurs d'exactitude, comme le pourcentage d'angles estimés avec une erreur inférieure à 5° (PCEA@5°) et l'erreur moyenne (MJAE), indiquent des écarts trop importants pour une utilisation clinique directe sans ajustement. L'application de corrections systématiques a amélioré ces performances, ramenant la MJAE moyenne à 11.2° et la RMSE à 15.6°, avec une PCEA@5° atteignant 46 %. Ces résultats montrent que de tels systèmes sans calibration peuvent approcher la viabilité clinique pour l'évaluation du StS dans des contextes contrôlés, mais nécessitent encore des améliorations pour une adoption plus large en clinique.

ABSTRACT

This thesis evaluates the clinical potential of low-cost, markerless motion capture systems for assessing sit-to-stand (StS) exercises in rehabilitation. Two off-the-shelf 3D cameras Intel RealSense D455 (stereo) and L515 (Time-of-Flight) were tested against a gold-standard Vicon system. A custom pipeline using the free software Nuitrack SDK was developed to extract joint angles and velocities from RGB-D recordings. The experiment was conducted with a single participant performing StS repetitions, captured simultaneously by both 3D cameras and the Vicon reference system. Precision was assessed by running the pose estimation process 30 times on each recording to evaluate the consistency of the results. Accuracy was then measured by comparing the estimated joint angles to the Vicon ground truth across 11 experimental conditions, designed to reflect realistic variations in execution speed, posture, symmetry, and clothing. Results show that the L515 yielded consistent trajectory shapes for several joint angles in standard StS scenarios, with Concordance Correlation Coefficients (CCC) averaging 0.91. However, other accuracy metrics, such as the Percentage of Correctly Estimated Angles within 5° (PCEA@5°) and Mean Joint Angle Error (MJAE), revealed discrepancies too large for direct clinical use. Applying systematic offset corrections improved performance, reducing average MJAE to 11.2°, RMSE to 15.6°, and increasing PCEA@5° to 46%. These findings suggest that while calibration-free systems like these are not yet ready for clinical deployment in their current form, they show promise for controlled environments and warrant further refinements for broader rehabilitation applications.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS, KEYWORDS AND ABBREVIATIONS

CCC	Concordance Correlation Coefficient
CI	Confidence interval
D455	Stereo 3D camera from Intel Realsense
dTOF	Direct time of flight
FPS	Frame per second
HPE	Human pose estimation
IMU	Inertial measurement units
ICC	Intraclass correlation coefficient
iTOF	Indirect time of flight
IR	Infrared
L515	Lidar 3D camera from Intel Realsense
LED	Light-emitting diode
Metrics	A quantifiable evaluation tool, referring to angles, positions, velocities and accelerations.
MJAE	Mean joint angle error
MPJPE	Mean per joint position error
OTS	Off the shelf
PCEA-$\epsilon°$	Concordance Correlation Coefficient and Percentage of Correctly Estimated Angles at $\epsilon°$
RGB-D	Red-Green-Blue-Depth
RMSD	Root mean standard deviation
RMSE	Root mean standard error
rpm	Repetitions per minute

StS	Sit-to-Stand
ToF	Time of flight

LIST OF APPENDICES

CHAPTER 1 INTRODUCTION

Lower limbs play an important role in daily mobility and autonomy. Walking, using stairs and standing up from a seated position are three important locomotor tasks that allow living in a wide range of environments. Losing these abilities can be due to a variety of reasons, ranging from traumatic accidents to aging and degenerative diseases. A wide array of medical specialists are trained to allow patients to recover and/or maintain these functions.

Assessing and treating mobility impairments requires specialists to closely analyze a patient’s movement and functional capabilities. Pictures and then videos were the first digital technologies used to enhance the data acquired from the sessions. Since then, a variety of motion capture systems have been developed in the goal of quantifying subtle differences in movement in a more reliable and valid way.

Marker-based motion capture systems have become an essential tool in clinical settings, providing a valuable complement to traditional, scale-based expert evaluations (Scataglini et al., 2024). These systems allow a single movement to be captured from multiple viewpoints and reanalyzed as needed. They enable the precise measurement of joint angles, velocities, and accelerations, facilitating detailed comparisons of patients’ performance over time.

Despite these advantages, the widespread adoption of marker-based systems has been limited by their high cost and time requirements (Mündermann et al., 2006); the setup, data collection, and post-processing can take several hours per patient, making routine clinical use impractical. Additionally, soft tissue artifacts can introduce errors in joint position estimates, leaving marker-based systems as the "silver standard". Biplane fluoroscopy, that uses X-ray to continuously image the bones is the “gold standard”, but the radiation dose, the small working volume and the even higher cost makes marker-based systems the most commonly used.

In contrast, markerless motion capture is a more recent technology that leverages advances in computer vision. It requires at the minimum a 2D camera but can also use multiple 2D cameras or a 3D camera. The state-of-the-art versions of these systems are powered by machine learning, making this technology particularly promising in this era of rapid AI advancements.

This master’s thesis introduces a proof of concept for the clinical use of “off-the-shelf” markerless 3D cameras for the evaluation and rehabilitation of sit-to-stand (StS). The system prioritizes affordability and convenience, using low-cost, portable cameras that require no calibration and are easy to deploy.

Such a system has the potential to significantly broaden the reach of motion capture technologies. Its low cost could make it accessible to more clinics, while its minimal setup time would allow for routine use in clinical workflows (Mündermann et al., 2006). Automated data processing and interpretation could enable stand-alone rehabilitation tools that provide real-time feedback to patients. Finally, its portability opens the possibility of evaluating mobility in a wide range of real-world settings, such as museums, parks, or shopping centres.

The next chapters of the document are organized as follows:

Chapter 2 presents a comprehensive background and literature review on StS rehabilitation as well as technology for digital analysis of movement. This survey provides the research basis of the thesis, as well as the current gaps in research, positioning this current thesis in the field of study.

Chapter 3 defines the research question and hypotheses that guide the study, along with the main goal and specific objectives.

Chapter 4 details the methodology, including the materials, software, experimental setup, and test conditions. It also outlines the data processing and analysis pipeline, with the aim of supporting replication and transparent interpretation.

Chapters 5 and 6 consist of the results and the discussion. These chapters give an overview of the results collected and analyze the performance and accuracy of the device under test. It also exposes the caveats of the system as well as the ethical considerations proper to the use of machine learning in health care.

Chapter 7 concludes this work with a summary of the main findings and their implication in the field of digital rehabilitation. Future works are proposed as well as avenues for related research.

CHAPTER 2 LITERATURE REVIEW

2.1 Background

Lower limb mobility is a key component of individual independence and autonomy during everyday activities. It allows for easy locomotion in a wide range of environments, especially those who have not been inclusively adapted to people with disabilities (Cardona et al., 2020). Fully functional lower limbs include the musculoskeletal features but also the proper functioning of the nervous system controlling and contributing to movement initiation, control, reflexes and somatosensory features.

2.1.1 Lower limb impairment

Conditions affecting mobility will be exposed in three categories: (i) trauma-related injuries, (ii) neurological diseases, and (iii) age-related limitations.

2.1.1.1 Trauma-related injuries

Traumas are sudden events coming from the external world injuring bodily tissues. Bone fractures are a type of trauma that might prevent mobility for a short period and require rehabilitation. Bones that can be immobilized tend to heal in a better way, thus reducing the risk of long-term impairment. The opposite is true for smaller bones such as in the ankle or the foot, or bones that are difficult to immobilize like the pelvis, which leads to a higher risk of walking impairment (Fernández-Gorgojo et al., 2022; Mahmood et al., 2024).

Vertebral column injuries due to broken or displaced vertebrae also affect mobility. Vertebrae fractures are not needed for debilitating symptoms to appear; soft tissues surrounding the column can be sufficiently damaged to affect the stability and trunk and lower body function. Long-term limitations can happen when the damage is too extensive and when spinal fusion is required. Spinal cord injuries can be caused by the traumas described above. Lesions have different severities and can sometimes heal through rehabilitation but may also lead to lasting impairments.

Amputations can follow a trauma and are likely to affect mobility. In such situations, prosthetics are often used and thus fall outside of the scope of this work.

Traumatic brain injuries, when brain damage is caused by an external impact, tends to affect balance and muscle control. Long-term multifaceted rehabilitation is often needed and supports patients in the recovery of their independence (Tramontano et al., 2022).

2.1.1.2 Diseases

Neurological diseases tend to affect mobility. Parkinson’s disease, cerebrovascular accident (CVA; stroke), arthritis and multiple sclerosis require continuous physiotherapy to improve comfort and fight the progression of the disease. In the case of amyotrophic lateral sclerosis (ALS) and cancer, which are sometimes terminal, the timeline can be shorter and rehabilitative practices are often used to improve the quality of life (Kasven-Gonzalez et al., 2010).

Cerebrovascular accidents are particularly prevalent and affect patients in various ways. Depending on the position and the size of the brain region affected, different impairments will show (Cardona et al., 2020). The first few months of rehabilitation post-stroke are crucial and indicate with great significance what level of disability will remain for the rest of the patient’s life (Cherry-Allen et al., 2023).

About 878,500 Canadian adults older than 20 years old have experienced a stroke and is a leading cause of long-term disability (Cherry-Allen et al., 2023; GovernmentOfCanada, 2023). Stroke is associated with a wide variety of motor deficits and somatosensory limitations (Cherry-Allen et al., 2023).

2.1.1.3 Age

Older populations have a much higher rate of morbidity affecting mobility (Campitelli et al., 2021). Progressive frailty and loss of balance make gait and StS more difficult for this population (Campitelli et al., 2021). When paired with osteopenia, the risk of traumas-causing-falls is increased, which on its own leads to a vicious cycle of comorbidity (Campitelli et al., 2021).

Physical therapy and activity are important for this part of the population that is generally more sedentary. These efforts should be made continuously and not exclusively after an adverse event. Since the decline is gradual, exercise has a big impact on the reduction of falls, the main component of the elderly vicious cycle (Zlatičanin et al., 2024). Hence, along with post-stroke individuals, this is one of the populations who could benefit the most from structured rehabilitation exercise.

2.1.2 Sit-to-Stand

The StS motion is defined as moving from a stable sitting position to a stable standing position. It is a primordial movement in daily life, such as leaving the bed, going to the bathroom and sitting up (Wang et al., 2022). The average person performs StS 60 times per day in order to complete daily tasks (Wang et al., 2022). Difficulty performing StS can reduce quality of life by adding both physical effort and mental strain. Tasks that are normally automatic may require conscious planning and execution, increasing the cognitive load with every movement. Such loads can lead to institutionalization in the elderly population, hence evaluation of StS is often performed in these populations (Janssen et al., 2002). Strength and power of the legs are needed when raising to a standing position, making it one of the most mechanically demanding functional daily activities (Boswell et al., 2023). It is associated positively with improved strength, balance and sensorimotor abilities (Campitelli et al., 2021). StS is also an indicator of well-being, kinematics of StS have been shown to be related to diagnosis of osteoarthritis, physical and mental health, Parkinson's disease, pulmonary disease, risk and history of fall, and degenerative spinal pathologies (Boswell et al., 2023; Masullo et al., 2019).

Disorders affecting StS, gait or balance affect 10% of people in their 60s and 60% of people over 80 years old (Cardona et al., 2020). When considering StS power, the average, peak and relative power value tend to decrease by around 10% per decade between the age of 50 and 80 years old (Campitelli et al., 2021). Given the current aging population and the mobility deficits in several patients' population, we can expect the prevalence of StS, gait or balance disorder to increase in the coming years.

2.1.2.1 Sit-to-Stand Biomechanics

StS is a movement requiring coordination of the whole body, with the primary joints involved being the ankles, knees, hips and trunk. The movement is often subdivided in four phases, the subdivisions from (Schenkman et al., 1990), seen in Figure 2.1, will be used in this work.

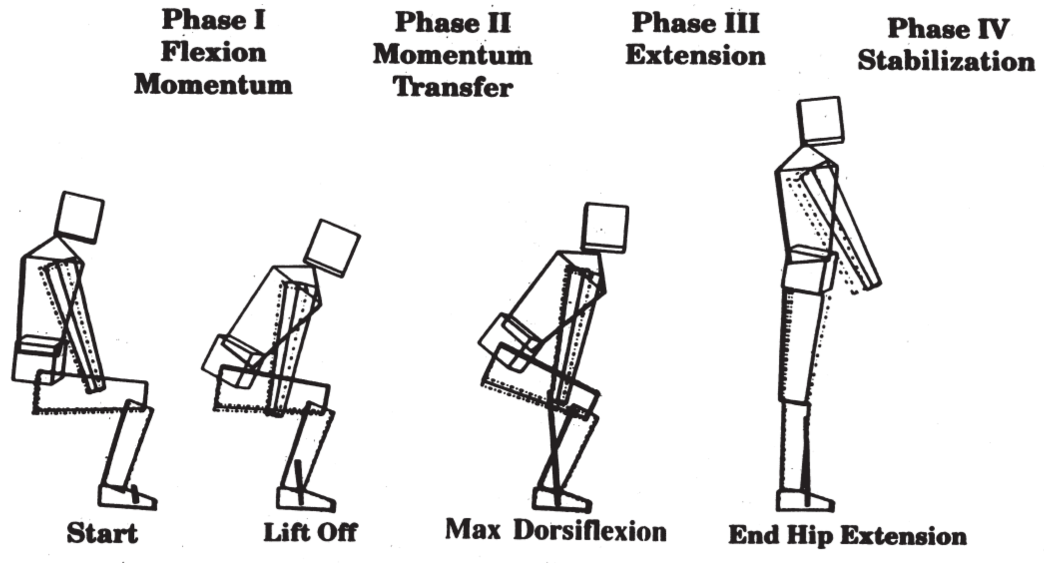


Figure 2.1 Diagram of the four phases of StS from (Schenkman et al., 1990).

- Flexion-Momentum Phase: This phase starts with a forward flexion of the pelvis and the trunk and ends just before seat lift-off. Forward momentum is generated by the upper body by bringing the trunk forward rapidly. The lower body is static during this phase since the hips are on the chair and the feet are on the ground (Janssen et al., 2002; Schenkman et al., 1990).
- Momentum-Transfer Phase: Starting at seat lift-off, the momentum is transferred to the rest of the body. Leaving the chair, the hips begin to move forward and upward, with the trunk remaining in forward flexion. Maximum ankle dorsiflexion ends this phase (Janssen et al., 2002; Schenkman et al., 1990).
- Extension Phase: Hips and knee extension begins in this phase, making the centre or mass move upwards rapidly. This phase ends when maximal hip extension is reached (Janssen et al., 2002; Schenkman et al., 1990). The most amount of force is required here, since the knee joints need to overcome a large torque (Wang et al., 2022).
- Stabilization Phase: Final stabilization happens in this phase and usually lasts less than a second in healthy individuals (Mao et al., 2018). The movements happening in this phase are of low amplitude and determining the end of the phase can be challenging (Mao et al., 2018; Schenkman et al., 1990).

These four phases describe a healthy momentum transfer strategy, but the momentum transfer phase requires good muscle control of the trunk and the lower limbs to generate enough forward momentum without later falling forward (Schenkman et al., 1990). A second strategy is common for patients without sufficient muscle control, which is called the Zero-Momentum Strategy. No momentum is generated with the trunk in this strategy and begins with a static trunk in flexion and feet placed under the centre of mass (Schenkman et al., 1990). The centre of mass moves only upwards with this strategy, lowering the amount of control and coordination required, but increasing the amount of force required in the legs (Schenkman et al., 1990). These two strategies present very different upper body velocities and range of motion, with the Zero-Momentum strategy requiring high trunk flexion, while the other one requires high trunk angular velocity (Schenkman et al., 1990).

2.1.3 Rehabilitation of the StS

StS rehabilitation constitutes just one component of a larger therapeutic strategy, and because each patient's needs differ, individualized programs are crucial. Detailed evaluations by physical therapists or other clinical professionals guide these programs, ensuring the chosen methods match the patient's current performance level. Importantly, all the disorders described in the previous section can impair a patient's ability to perform StS and may therefore require targeted rehabilitation. Because StS is a fundamental activity of daily living, restoring or improving this function is often a meaningful goal in itself. However, the decision to include specific StS exercises or evaluations in a rehabilitation plan depends on the patient's functional baseline and goals (Chaovalit et al., 2020).

Evaluation by clinical staff such as physical therapists is the most common evaluation method since it is the most seamless. Timed evaluation, such as the 5-time StS or the 30 seconds StS test, is an objective measure associating the time spent or the number of repetitions with a certain level of performance. The 5StS is often used clinically to evaluate the ability of the patient to rise and sit multiple times at a rapid rate. The resulting time can be found on a chart based on age and helps identify a risk of fall or disorders ("Five Times Sit to Stand (5TSTS)," 2019).

The score can be used to infer on the training regimen and can also be used for eventual comparison through time with the same patient to see improvement. This objective method strips away most of the information other than movement speed, hence qualitative description can accompany the timed evaluation result such as stability and movement smoothness.

To bridge this gap, motion capture systems have been introduced in rehabilitation science, offering detailed kinematic and dynamic data for diagnosis and treatment planning. Engineers and clinicians strive to enhance user-friendliness, accuracy, and accessibility of motion capture, aiming to make it a practical standard of care.

2.1.4 Digital Analysis of Movement

Multiple technologies exist to digitalize movement and locomotion, each being more suited to a certain application. No system exists that combine accuracy, ease of use and cost at the highest level, therefore different technologies are being used simultaneously and in different settings.

2.1.4.1 Inertial Systems

Inertial Measurement Units (IMUs) are small integrated sensors containing a gyroscope and accelerometers to measure linear and angular accelerations, and sometimes magnetometers to orient the device according to an external magnetic field. IMUs are secured to the subject at specific anatomical locations. They do not require line of sight, are portable and can have a long reception range, making them useful in large outdoor environments (Cardona et al., 2020). Their accuracy is limited by drift over time, which is caused by the problem of double integration of the acceleration, an inherently noisy process. (Cardona et al., 2020).

Figure 2.2 Example of a commercial inertial measurement unit system: the Movella Xsens Awinds (Movella)

2.1.4.2 Goniometers

Planar joint angle measurement can be measured with goniometers. They are electromechanical instruments attached to the body on both sides of the joint, in the axis of rotation wanted. They are inexpensive and give repeatable results when calibrated accurately (Akhtaruzzaman et al., 2016).

Although, they have been shown to work better with the elbow joint than the knee or joints with more than one degree of freedom (Akhtaruzzaman et al., 2016).

Figure 2.3 Example of a wireless commercial goniometer from Biometrics LTD (BiometricsLTD, 2024)

2.1.4.3 BiPlane Fluoroscopy

This technique uses X-ray imaging for dynamic tomography of a moving subject. Although a single X-ray source can be used, biplane fluoroscopy employs two sources to triangulate 3D positions. While it avoids the soft-tissue artifacts seen in marker-based systems and IMUs, it is costly to operate, has limited applications because of its generally small working volume, and it can increase the risk of cancer due to its delivery of a relatively high X-ray dose.

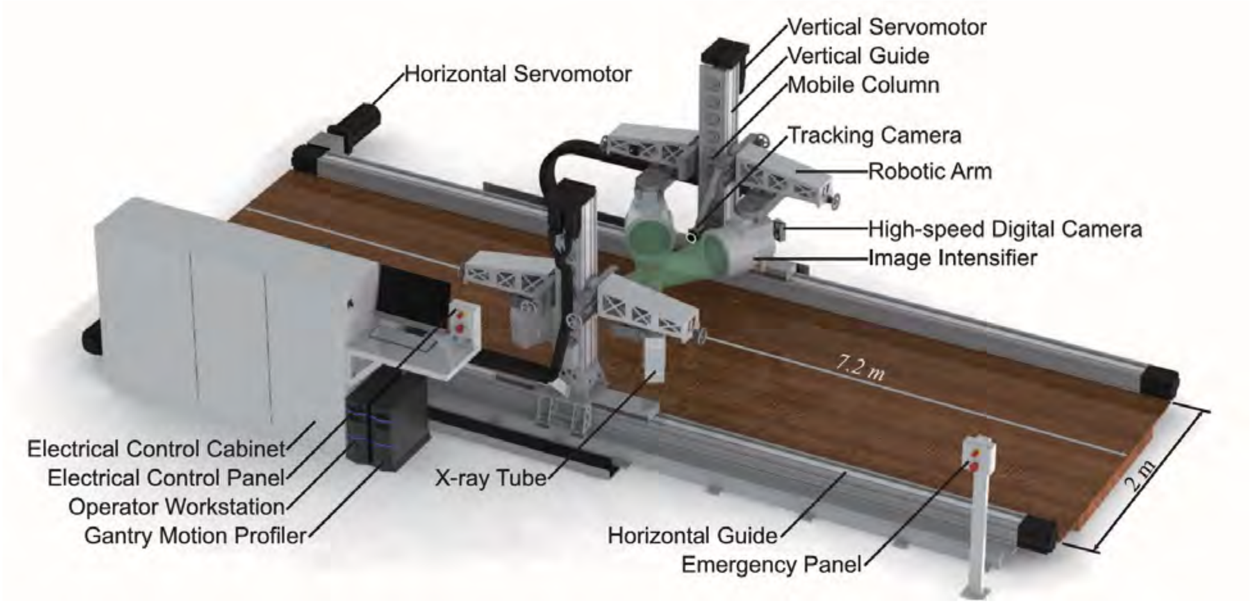


Figure 2.4 Example of a mobile biplane fluoroscopy system as used in (Guan et al., 2016)

2.1.4.4 Electromyography and Force Plates

These two technologies cannot give 3D position of joints but instead informs about invisible physiological information. They are often used in concert with IMUs or marker-based systems, offering another dimension to the movement analysis. Electromyography uses electrodes attached to the skin to read the intensity of muscle activation. Force plates measure forces and torques applied by the subject on the environment, closing the loop for dynamic analysis.

Figure 2.5 Example of commercial force plates developed by Bertec Corporation (Bertec, 2025)

2.1.4.5 Marker-Based Motion Capture Systems

Marker based systems rely on physical markers to be placed at specific anatomical places on the skin of the patient. An array of cameras surrounds the subject, capturing the position of the markers in their 2D field of view. When a marker is seen by two cameras simultaneously, the 3D position of that marker can be inferred using triangulation. When a marker is seen by more than two cameras, the data is aggregated leading to lower noise and better precision of the 3D values.

Passive infrared markers are normally coated in a reflective material to appear bright under the infrared flashes of active systems. Active markers, in contrast, emit light allowing for individual identification of markers and higher signal-to-noise ratio, at the cost of higher technological complexity (Cardona et al., 2020).

Figure 2.6 Example of a commercial marker-based camera, the Vicon Valkyrie (Vicon, 2025b)

2.1.4.6 Markerless Pose Estimation

This technology relies on software to estimate the position of joint centres. These estimations are either rule based (Glandon et al., 2019) or based on machine learning models (Cong et al., 2022; Fürst et al., 2020). In the case of 2D pose estimation, any video can be the input to the model and the best estimation of the joint is attributed to a pixel of the video at every frame (Horsak et al., 2024). Relative positions and angles can be obtained from 2D pose estimation, but absolute measurements require more assumptions and processing, due to the absence of depth information and scale ambiguity inherent in 2D imaging (Scataglini et al., 2024).

All 3D cameras can be used to solve this ambiguity and offer valuable information to the human pose estimation (HPE) algorithm. While markerless systems are not as precise as their more established counterparts, they benefit from being less intrusive, less time intensive and less expensive (Avogaro et al., 2023). More information on the hardware used and on HPE algorithms can be found in the following sections.

2.1.5 Movement Evaluation Metrics

Quantitative evaluation metrics provide a structured way to evaluate the quality of StS movements and highlight areas of impairment or improvement. The metrics can be broadly categorized into (i) temporal, (ii) kinematic, (iii) kinetic, and (iv) muscle activity parameters. Some measurements can be used as it is, while other metrics require further processing; for example, when comparing the right and left side of the body (Janssen et al., 2002).

2.1.5.1 Temporal Metrics

Temporal metrics have units of seconds or repetition/second. The evaluation is often for five repetitions or for the most repetitions in a 30s period (Boswell et al., 2023). They can also evaluate a certain phase of the StS motion by measuring the time to seat lift-off for example (Janssen et al., 2002; Mao et al., 2018). These metrics can be obtained with a timer in real time and mostly do not require motion capture.

2.1.5.2 Kinematic Metrics

Kinematics metrics are based on positions, velocities or accelerations, linear and angular (Cardona et al., 2020; Hyodo et al., 2023). The angles of the knee, hip, ankle in the sagittal plane are of particular interest as well as the trunk angle in the sagittal and frontal planes (Bhardwaj et al., 2021; Boswell et al., 2023; Hyodo et al., 2023). The range of motion, maximum, minimum and means of joint angles during the task can be calculated according to the need and limitations of the patients being evaluated (Hyodo et al., 2023). Arm position is also described, although they are often crossed or unused during clinical StS evaluation.

2.1.5.3 Kinetic and Muscle Activity Metrics

Kinetic metrics are based on forces and torques. Force plates are required alongside kinematic measurements for accurate kinetic computations, but estimates of certain forces or centre of mass can be calculated from motion capture if certain assumptions are made. Force sensors on the chair can help with the timing of phases, such as loss of contact with the chair. Ground reaction forces are used in conjunction with inverse kinematics to find the centre of mass, joint moments and the power generated at specific joints (Campitelli et al., 2021; Cardona et al., 2020; Hyodo et al., 2023; Mao et al., 2018; Wang et al., 2022).

2.1.6 Hardware Technology Used for Human Pose Estimation

Generally, more precise and accurate data coming from the sensor will result in better HPE performance. This is the reason why commercial marker-based systems require an array of cameras with precisely positioned markers. However, the type of system studied in this thesis cannot afford these advantages. A larger system, in the form of a large baseline or large sensors, only improves accuracy at a cost of a higher retail price and less portability.

Two technologies of sensors are currently the most used for single-camera HPE: stereo and Time of Flight (ToF). They both solve the depth ambiguity, a problem inherent to a single monocular camera system, using two different techniques, as described in the following sections.

2.1.6.1 Stereo and Structured Light Cameras

Stereo cameras come in two kinds: passive and active. Passive stereo cameras use two monocular cameras at a set baseline distance from one another to solve the depth ambiguity problem. Edges, colours and textures are used to match features of the scene to a pixel position at each frame coming from both cameras. Depth can be inferred since objects closer to the lens will have a larger pixel disparity than objects far away. The general formula is shown below, where Z in the depth, f is the focal length, B is the baseline distance between the cameras, and d is the disparity of position between both images in pixels.

$$Z = \frac{f \cdot B}{d} \tag{2.1}$$

A large baseline distance will increase the depth sensitivity of the stereo camera but will increase the size of the device. Similarly, an increased focal length will also improve depth sensitivity but at the cost of a smaller field of view.

Structured light cameras use a projector in addition to one or two monocular cameras. The projector shines structured light on the scene overlaying the scene with varying engineered patterns, such as a grid, a speckle field or fringes. If a single monocular sensor is used, pattern deformation, occlusion, size and shapes of the structured light will inform the depth detection. The patterns projected are complex and vary in time to remove ambiguity. Multiple patterns per frame are needed, lowering the framerate. If two monocular sensors are used along with the projector, the structured light is used to enhance the scene by projecting artificial textures to low-feature surfaces (Xu et al., 2021). In this case, the structured light pattern can be static, and a higher framerate is achievable.

Figure 2.7 An example of a structured light method called the speckle pattern as used in the Kinect V1. It is a pseudo-random pattern of dots that helps for pattern matching to align features found in both images. (Clark et al., 2019)

The structured light is generally infrared (IR), which makes it invisible to the eye of the subject being illuminated as well as to the RGB sensor. There is also less interference coming from ambient visible light such as LEDs or neon lights which mostly emit visible light with very little IR leakage (Kumar et al., 2018). Conversely, sunlight and incandescent bulbs do emit IR light, which can interfere with the projection and make these IR-based cameras less performant in sunlight or under strong incandescent bulbs, but very advantageous in low light conditions or when the scene has little contrast.

2.1.6.2 Time of Flight Cameras

Time of Flight (ToF) is the second type of depth-sensing technology most available today. This method relies on calculating the round-trip time of the light shone onto a scene. Two technologies of ToF exist: direct and indirect.

Direct ToF (dToF) works by sending individual beams of light using lasers or LEDs and timing the round-trip travel of the light. They can be of the “scanning” form where mirrors are used to direct the beam on the scene at one point at a time, or of the “flash” form where the whole scene is illuminated at once and a 2D sensor array captures the return time per pixel. The distance from the camera, Z, is found using the following formula, where c is the speed of light in air and t is the time difference between emission and reception (Clark et al., 2019).

$$Z = 0.5\, t \cdot c \tag{2.2}$$

Indirect ToF (iToF) works by shining a continuous waveform of IR light at the entire scene. The waveform is modulated at specific frequency and the phase shift of the light coming back infers the distance from the target. Just like for AM and FM radio, both amplitude and frequency can be modulated, but with ToF, amplitude modulation is more common. The frequency of the modulation generally ranges between 10 MHz and 150 MHz. This modulation frequency is critical to the precision of the depth sensing, since a phase shift of 2π is indistinguishable from a phase shift of zero. This fact marks a limit to the maximal range that can be detected by the device. This distance is calculated with the formula $d_{max} = 0.5c \cdot f^{-1}$, where d_{max} is the maximum unambiguous distance, c is the speed of light and f is the modulation frequency. A modulation frequency of 10 MHz gives a maximum distance of 15 metres, and 150 MHz gives a maximum distance of one metre. The distance Z to the object can be calculated using the formula below, where c is the speed of light, ϕ is the phase shift and f is the modulation frequency.

$$Z = \frac{c \cdot \phi}{4\pi \cdot f} \tag{2.3}$$

This equation shows us that for a given $\Delta\Phi$ calculated by the device, a higher modulation frequency will lead to a smaller ΔZ, proportionally to the range.

2.1.6.3 Real world conditions and use cases

In rehabilitation, the digital evaluation of StS, gait or stability is usually done in a lab environment because of the limitations inherent to the motion tracking system. A wider range of real-world scenarios have the potential to be evaluated when using 3D cameras instead of marker-based systems.

Being aware of these different technologies allows us to make better choices when it comes to selecting a device for a specific scenario. For example, evaluating StS in a museum has different requirements than evaluating gait in a park environment. In Table 2.1 below, multiple scenarios are presented, and the associated suitability of each type of camera is given. The results are classified as good performance (+), average performance (0) and poor performance (-).

Table 2.1 Scenario-based comparison of the four main types of depth-sensing cameras.

Scenario		Passive Stereo	Active Stereo	Direct ToF	Indirect ToF
Light environment	Bright sunlight	+	-	0	-
	Normal light	0	+	+	+
	Low light	-	0	+	+
Depth range needed	Deep	0	-	+	-
	Shallow	0	+	0	+
Movement speed	Fast	0	-	0	0
	Slow	0	+	+	+
Resolution in 2D	N/A	+	+	-	0
Resolution in depth	N/A	0	+	+	0
Budget needed	N/A	+	0	-	-

For light environment, a system needing to shine uncollimated IR light will be susceptible to IR interference from strong rays from the sun or incandescent bulbs. Conversely, in low light environment, when the scene has low contrast, the IR light becomes an asset, making the system more performant (Xu et al., 2021). dToF systems that use collimated or focus light are much more robust to IR interference (Cong et al., 2022). Some can be used outdoors, depending on the brightness of the laser - a characteristic often associated with the price of the device (Cong et al., 2022).

When scenes require large subject displacements or extended tracking ranges, a greater depth range becomes essential. In active stereo and iToF systems, the intensity of projected infrared light diminishes rapidly with distance due to beam divergence, reducing signal quality and degrading depth accuracy. Furthermore, as the subject moves farther from the sensor, fewer pixels cover each physical area of the scene, inherently lowering spatial precision. These combined effects make active stereo and iToF less suitable for long-range applications. In contrast, dToF systems typically employ collimated laser pulses that maintain high brightness over greater distances. While spatial resolution reduces just as quickly as iToF, dTof does not suffer from the additional beam divergence, making it more suitable to long-range applications.

When the movement under study is fast, the framerate and motion blur become an issue. A low framerate can result in under-sampling of the motion, missing key features or transitions. A long exposure time causes motion blur, which can obscure fine details and reduce the accuracy of HPE. The RGB sensors from stereo cameras can operate at a high framerate with little exposure in well-lit environments, but not all devices have this capacity (Akhtaruzzaman et al., 2016). Low framerate stereo devices will perform poorly with fast movements, since motion blur will fade the features required for pattern matching. Active stereo and iToF with a single monocular sensor have the limitation of needing multiple pulses of light at every frame, limiting their maximum framerate and needing a long exposure time. The dToF sensors only need a single pulse of light per pixel, making the image very sharp, but the raster scanning of the laser through the scene can cause a rolling shutter effect.

2.1.6.4 Other System Types used for HPE

Beyond stereo and ToF sensors, HPE can use two or more calibrated monocular cameras. Viewing the subject from multiple angles allows triangulation of joint positions with higher accuracy and less sensitivity to occlusion. Adding cameras generally improves robustness. At the high end, systems like Theia3D use several synchronized high-resolution cameras with advanced model-based tracking, achieving markerless accuracy close to marker-based systems(TheiaMarkerless, 2025). However, each extra camera increases the need for precise calibration and tight synchronization. This makes such setups less portable and more complex to deploy.

2.1.7 Software Technology Used for HPE

The second layer of technology for HPE is software: it is the system that extracts joint positions from visual data. The output can be 2D or 3D, irrespective of the input data, since depth ambiguity can be reduced with assumptions.

There are two core approaches to HPE: top-down and bottom-up. The top-down approach starts by identifying one or multiple people in the image and then identifies their joints. This method works poorly in high occlusion scenarios, where a subject could be missed altogether, but is more robust to false positives (Avogaro et al., 2023). The bottom-up approach starts by finding body joints within the image and then associating the joints with each person. In situations where body parts are occluded, bottom-up works well, but it also tends to falsely identify parts of the scene as body parts (Wang et al., 2022).

2.1.7.1 Software Architecture

The field of HPE has been dominated for many years by rigid, rule-based, template-driven algorithms (Glandon et al., 2019). These algorithms are computationally efficient, but are highly dependent on the coded rules, breaking as soon as the situation falls outside of what is expected. Machine learning has increasingly been used to replace the rule-based architecture, replacing the need for robust rules, by the need for big data.

Convolutional neural networks are the current state of the art for real-time HPE. They work by extracting local spatial features and extracting meaning out of them using a long training process. They are computationally expensive but can be used in real time.

Transformer based techniques are getting more popular for complex scenes because of their large-scale spatial and temporal dependencies. This unfortunately comes at a cost: the complexity scales quadratically with input length, making them prohibitively costly for real-time tasks using consumer-level hardware (Selva et al., 2022).

2.1.7.2 Temporal Models

Video based HPE can be seen as a series of picture-based HPE, but doing so would neglect valuable information. Temporal information is used to give consistency across frames and reduce high-frequency noise. It can be applied as a filter after the joint positions are obtained, or it can be built directly into the model. While post-filtering methods are less computationally expensive, integrating temporal reasoning directly into the model, such as with transformer-based architectures, offers a more robust approach, enabling the system to infer occluded joints over time, much like the human visual system (Selva et al., 2022).

2.1.7.3 Parametric 3D Body Models

An alternative to skeletal-only HPE is to fit a parametric 3D body model such as SMPL (Skinned Multi-Person Linear) directly to RGB or RGB-D images. SMPL jointly models pose and body shape, producing a realistic surface mesh from which limb volumes, body segment masses, and more accurate centre-of-mass estimates can be derived. Because the model captures the body’s outer surface, it holds the promise of better accounting for clothing and occlusions than joint-only methods, leading to more stable kinematics in challenging conditions.

2.1.7.4 Limitations of Machine Learning Based HPE

Any machine learning algorithm requires data to be trained effectively. This creates a new type of limitation, where the best models are often the ones trained on the most data. For proper training, the data needs to contain both the input of the camera and the desired output, which creates two types of limitations.

First, the input is very sensitive to any type of distortion coming from the camera and works best if the camera used for the training is the same as the one used for testing (Yang et al., 2021). Robust algorithms can be trained to compensate for the various cameras, but the characteristics of each camera needs to be learned with even more data (Yang et al., 2021)

The second limitation is the need for a ground-truth output. Labelling large datasets requires a lot of effort and resources. Microsoft has published the largest dataset yet, used extensively in machine learning. The COCO dataset contains annotations of body parts, objects and descriptions that were generated manually by humans (Lin et al., 2014). Notably, the most widely used dataset was developed by one of the largest technological companies today: Microsoft (Yang et al., 2021).

For these two reasons, the availability of large, labelled datasets has triggered a gold rush effect, where certain subsets of data receive disproportionate attention simply because they are readily accessible. This bias toward convenience often shapes research focus more than clinical relevance. These limitations will be further examined in the discussion section.

2.2 Literature Review

Markerless motion capture using 2D and 3D cameras has become a rapidly advancing area of research, driven by several intersecting technological trends. Notably, recent breakthroughs in neural networks have led to the development of high-performing HPE algorithms, such as OpenPose and various proprietary systems (Martınez, 2019). At the same time, progress in fields like autonomous driving has pushed markerless pose estimation to the forefront of computer vision research, resulting in more accurate models and the accumulation of vast, diverse datasets for training. These combined improvements have enabled the medical and rehabilitation communities to begin applying markerless pose estimation for clinical tasks such as diagnosis, treatment planning, and patient monitoring. In this section, research using markerless pose estimation will be presented.

2.2.1 2D Pose Estimation from a Monocular Camera

HPE in two dimensions (2D-HPE) is the first technological stone on top of which all other techniques are built. It consists in extracting from an image the location of body landmarks for every frame. OpenPose is a 2D-HPE machine learning algorithm based on convolutional neural networks that are highly performing and widely used (Boswell et al., 2023; Cao et al., 2018; Martınez, 2019). While it is not the only open-source application that can extract kinematic data from a video, it is generally considered the best (Washabaugh et al., 2022).

The output of 2D-HPE does not have any scale so the information can only be relative. In a paper from (Boswell et al., 2023), they use OpenPose to extract parameters of StS, such as the timing, knee and trunk angles, speeds and accelerations. The subjects were viewed at a 45-degree angle, which allows sagittal angles to be observed while keeping occlusions to a minimum. The parameters gathered could predict a diagnosis of osteoarthritis and estimate physical and mental health among other individual characteristics. Subjects filmed their video on their own and most considered the process of filming and uploading the video very easy, opening the door to a new form of large-scale diagnostic of groups (Boswell et al., 2023).

Figure 2.8 Use of OpenPose with a smartphone camera in (Boswell et al., 2023) for StS evaluation.

The main limitation of this form of HPE is its inability to capture anything in the depth axis. Using a 45-degree angle for the video, a movement of the trunk in the sagittal plane or in the frontal planes gives the same result on the video. This ambiguity requires multiple assumptions to be made, for example, that all movements are along the sagittal plane. Conversely, the main advantage of this method is that it does not require a specialized camera, lowering the barrier of entry.

2.2.2 3D Pose Estimation from Two Monocular Cameras

3D-HPE using two monocular cameras is the natural next since it consists of merging two 2D-HPE systems. A good calibration is required to merge the two cameras, both in time and in space. Calibration in space must span the whole working volume to account for distortions. This technology generally does not need the assumption required for 2D-HPE, since occlusions are less problematic.

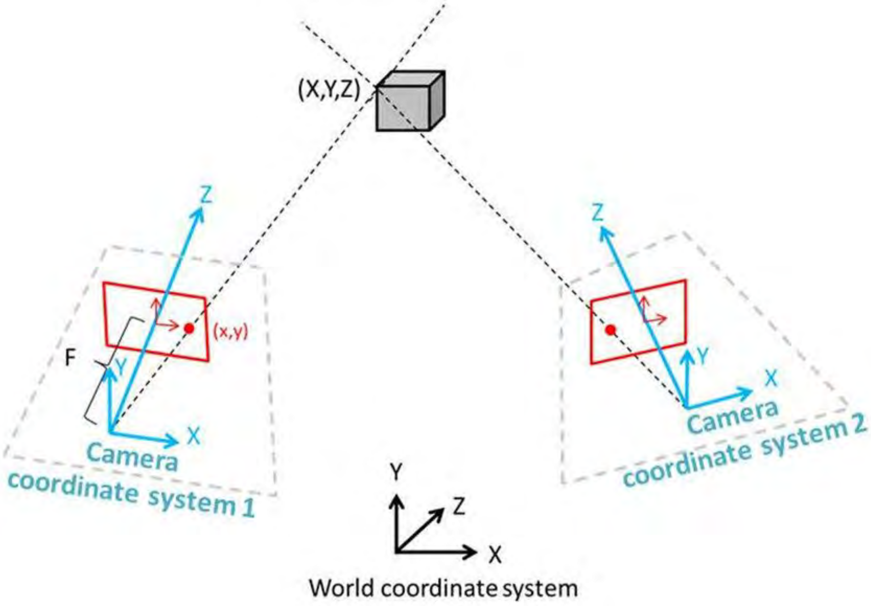


Figure 2.9 Diagram of stereoscopic triangulation using a two-camera system (Ekberg et al., 2017).

Two studies used a pair of monocular cameras placed on either side of the body to obtain sagittal angles without occlusions. OpenPose was used again for both studies. (Serrancolí et al., 2020) used a pedalling setup, while (Stenum et al., 2021) used a walking setup, but in both cases, each camera extracted information from the side of the body closest to it. (Serrancolí et al., 2020) compared the estimated sagittal joint angle to the joint angles identified manually from the video. The root mean square error (RMSE) of the angles were less than 3° for the hip up to less than 11.5° for the ankle. (Stenum et al., 2021) compared the estimated angles to Vicon data and found a mean absolute error between 4.0° and 7.4° for the hip, knee and ankles. Both studies concluded that ankle angle was not sufficiently precise to be used clinically, while knee and hip angles were accurate and precise enough to detect a significant change.

(Horsak et al., 2024) used a two-camera system and OpenCap (Uhlrich et al., 2023) for the analysis of gait, where the two smartphone cameras were placed symmetrically, 35° relative to the direction of walking. Vicon was used as the ground truth with OpenSim Muskuloskeletal Modelling (Seth et al., 2018) to extract the angles from Vicon and OpenCap. They compared gait cycle consistency for multiple pathological gait patterns and showed that the markerless variability was much higher than the ground truth variability. Pelvis tilt and hip flexion were the two parameters that contributed most to variability in measurements. The various gait patterns tested showed no significant difference in accuracy.

3D-HPE using two monocular cameras achieves much better results than 2D-HPE and can be used for a wide range of applications. The main limitation of this system is the reliance on spatial and temporal calibration of the cameras. This time-consuming step must be done on site and is vulnerable to any movement of the cameras.

2.2.3 3D Pose Estimation from a Single 3D Camera

Pre-calibrated dual-camera systems exist in the form of a stereo camera. The Microsoft Kinect was the first mass-produced inexpensive 3D stereo camera, making it widely available. It was associated with the Xbox gaming system, opening the door to new types of games using the body as a remote control. The Kinect V1 is an active stereo camera using structured infrared light to gauge depth and the Kinect V2 uses a ToF method, similar to lidar technology but without the use of a laser. Even though the Kinect V2 is more recent, it does not always perform better than the V1 for depth map accuracy and skeleton tracking. There are also differences within versions, making it particularly difficult to make version-dependent conclusions (Clark et al., 2019).

The Kinect has been used extensively in rehabilitation studies, as discussed in the (Clark et al., 2019) review article. Studies have evaluated the accuracy of balance, posture, gait, StS, stand-up and go, kinematic analysis and inverse dynamics. Overall movement of large body parts such as the trunk can be measured accurately but small changes in positions (< 2 cm) are hidden by noise. Some spatiotemporal variables, such as step length and step width, were also found to be valid for gait analysis. In contrast, parameters from smaller body parts, such as the knee angle, show larger errors that cannot be used clinically. Torque data extracted using inverse kinematics is also unreliable, having an error above 20% for most of the variables. (Dolatabadi et al., 2016) found that the Kinect could be used to distinguish between healthy and pathological gait using machine learning, but it is unclear how this performance compares to a visual evaluation (Dolatabadi et al., 2016). In general, the trend of the joint trajectory is accurate, but magnitudes can have large errors and cannot be used reliably.

Apple introduced a lidar camera in its pro range of the iPhone and iPad starting with the 12th generation, increasing the ubiquity of lidar sensors. In the paper from (Kaewrat et al., 2024), an iPhone 13 Pro is used to evaluate a use case of telemedicine in multiple room conditions, and subject body mass index. Although they did not have a ground truth in the form of marker-based motion capture, they did show that lidar-based tracking outperformed RGB-based tracking in fault detection accuracy across various environments, particularly in complex or reflective backgrounds, while RGB cameras scored higher on usability due to their screen-facing placement and visual feedback. Other RGB-D lidar and stereo cameras exist at a wide price range, but rare are the studies evaluating their precision and accuracy in a rehabilitation environment.

2.3 Summary and Project Rationale

While the field of HPE has advanced significantly in recent years, particularly through the development of a neural network–based methods and the widespread availability of depth-sensing hardware, important gaps remain. Many studies have relied on 2D-HPE models or stereo setups that require careful calibration, making them less practical in uncontrolled or clinical environments. Studies using 3D cameras have been performed with very low-cost devices such as the Kinect V1, Kinect V2 or the iPhone lidar and have mostly concluded that their accuracy was not sufficient for clinical use. In cases where more expensive systems were used, the objectives were multi-person HPE in the context of autonomous driving or crowd surveying, where the accuracy and reliability of small joint movements or fine-grained clinical assessments remain underexplored. Most importantly, there is a lack of quantitative validation studies examining low-cost, markerless systems specifically in the context of StS rehabilitation, a movement more suited to markerless HPE than gait, while being critical for maintaining autonomy in aging and mobility-impaired populations.

These gaps highlight the need for an affordable, easy-to-deploy HPE system that can operate without calibration in an uncontrolled environment, maintain reasonable accuracy, and be evaluated for clinical relevance. This thesis addresses that need by proposing and testing a proof-of-concept pipeline using off-the-shelf 3D cameras and open-source machine learning–based pose estimation in a StS rehabilitation context.

CHAPTER 3 RESEARCH OBJECTIVES

3.1 Research Question

How accurately and reliably does a cost-effective markerless motion capture system capture key kinematic parameters pertinent to StS exercises compared to traditional motion capture methods?

3.2 Hypotheses

Hypothesis #1: Single-device, plug-and-play 3D cameras offer the versatility and ease of use required to make motion capture practical and efficient in clinical settings.

Hypothesis #2: Standard StS tasks can be evaluated with markerless HPE and yield measurements that are sufficiently accurate and precise for clinical evaluation.

Hypothesis #3: StS movements reflecting the range of patterns seen in rehabilitation, including compensatory strategies and clothing-related variations, can also be evaluated using markerless HPE, provided appropriate post-processing steps are applied.

3.3 Main objective and specific objectives

The main objective of this study is to assess the clinical validity of a low-cost, markerless motion capture system by comparing its performance against a commercial reference system, with the goal of expanding access to motion-tracking technology for a broader range of patients in need.

To achieve the main objective, the following specific objectives are defined:

- SO1: Identify the practical and technical requirements for deploying markerless motion capture systems in clinical environments and assemble a low-cost calibration-less HPE system using off-the-shelf hardware and software.
- SO2: Evaluate the system’s performance across multiple kinematic metrics, comparing results to a gold-standard motion capture system in order to identify which metrics remain robust under lower-cost conditions.
- SO3: Evaluate the system’s performance for multiple conditions and use case scenarios to identify limitations related to motion, asymmetry or clothing.

CHAPTER 4 METHODOLOGY

This chapter outlines the methodology used in this research. It begins with a presentation of the three vision systems and the software associated with each camera. The experimental protocol is then described, including the tasks performed, the setup of the environment, and participant details. Next, the tools and procedures used to process the data from both the 3D cameras and the Vicon system are introduced. Finally, the chapter details the data analysis steps, including preprocessing, metric calculations, and methods for comparing the results.

4.1 Material

4.1.1 Desktop 3D Cameras

Two 3D cameras are evaluated in this study: the Intel RealSense Depth Camera D455 (Figure 4.1) and the Intel RealSense LiDAR Camera L515 (Figure 4.2. Both are compact devices, each weighing around 100 g and able to fit comfortably in a single hand. They require to be plugged into a computer for power, control and data management. The cameras do not require calibration and include an IMU to adjust the 3D scene with respect to gravity.

Figure 4.1 Intel RealSense Depth Camera D455

A depth image and an RGB image are captured by the camera and the alignment of both streams gives an estimate of the distance between the camera and the pixels of the RGB image. In Table 4.1, technical specifications of the D455 and L515 are displayed.

Figure 4.2 Intel RealSense LiDAR Camera L515

Table 4.1 Technical Specifications of the 3D Cameras

	D455 (IntelRealSense, 2025a)	L515 (IntelRealSense, 2025b)
Technology	Stereoscopic	ToF
Dimensions	124 mm × 26 mm × 29 mm	61 mm diameter × 26 mm height
Ideal Range	0.6 m to 6 m	0.25 m to 9 m
Depth accuracy	<2% at 4 m	5 mm to 14 mm at 9 m
Framerate	30 fps	30 fps
Resolution	1280 × 800 (RGB) Up to 1280 × 720 (depth)	1920 x 1080 (RGB) 1024 x 768 (depth)
Field of view	90° × 65° (RGB) 87° × 58° (depth)	70° × 43° (±3°) (RGB) 70° × 55° (±3°) (depth)
Price (as of 2025)	570 CAD	801 CAD

4.1.2 Vicon Cameras

The Vicon Cameras were mounted to the ceiling of a rehabilitation laboratory. Nine Vicon Vantage V5 cameras were surrounding the scene. These cameras have CMOS sensor with a resolution of 2432x2048, an 8-bit grey scale and are sensitive to 850 nm infrared light. The illumination is supplied by a strobing ring light consisting of 22 adjustable 850 nm LEDs. For this experiment, the framerate was set to 120 Hz.

For logistical reasons, data collection was conducted with the subject off centred with respect to the cameras, limiting the view of the markers on the left lower body in certain positions.

4.2 Data Collection Software

The material was connected to computers and controlled using their native software. The following section details the software used and the settings selected.

4.2.1 Intel Software

The application Intel RealSense Viewer version 2.49.0 was used in conjunction with the D455 and L515. This application is available free of charge as part of the Intel RealSense SDK. Two computers were used in parallel, with a single camera connected to each, to reduce the computational load. The camera settings were chosen to be “High Accuracy” on the D455 and “Max Range” on the L515, after some initial testing with all the different modes in the same environment. All data streams were enabled: the RGB data, depth data and orientation data.

4.2.2 Vicon Software

The Vicon Nexus software was operated by a professional. Calibration of the scene was done using a Vicon Passive Wand. Calibration of the subject with the markers was done statically and then dynamically, as prompted by Vicon Nexus.

4.3 Experimental setup

4.3.1 Room Setup

The setup of the StS experiment was as follows. The experiment room consisted of a standard 42cm tall chair positioned 2.6 m ahead of a table in a well-lit laboratory. The two 3D cameras were placed on the table at a standard counter height of 36 in (91 cm). The pitch of the cameras was adjusted to keep the subject in frame for the sequence of movement. The subject was facing the 3D cameras for all StS repetitions to avoid obstruction of body parts. The Vicon cameras surrounded the scene.

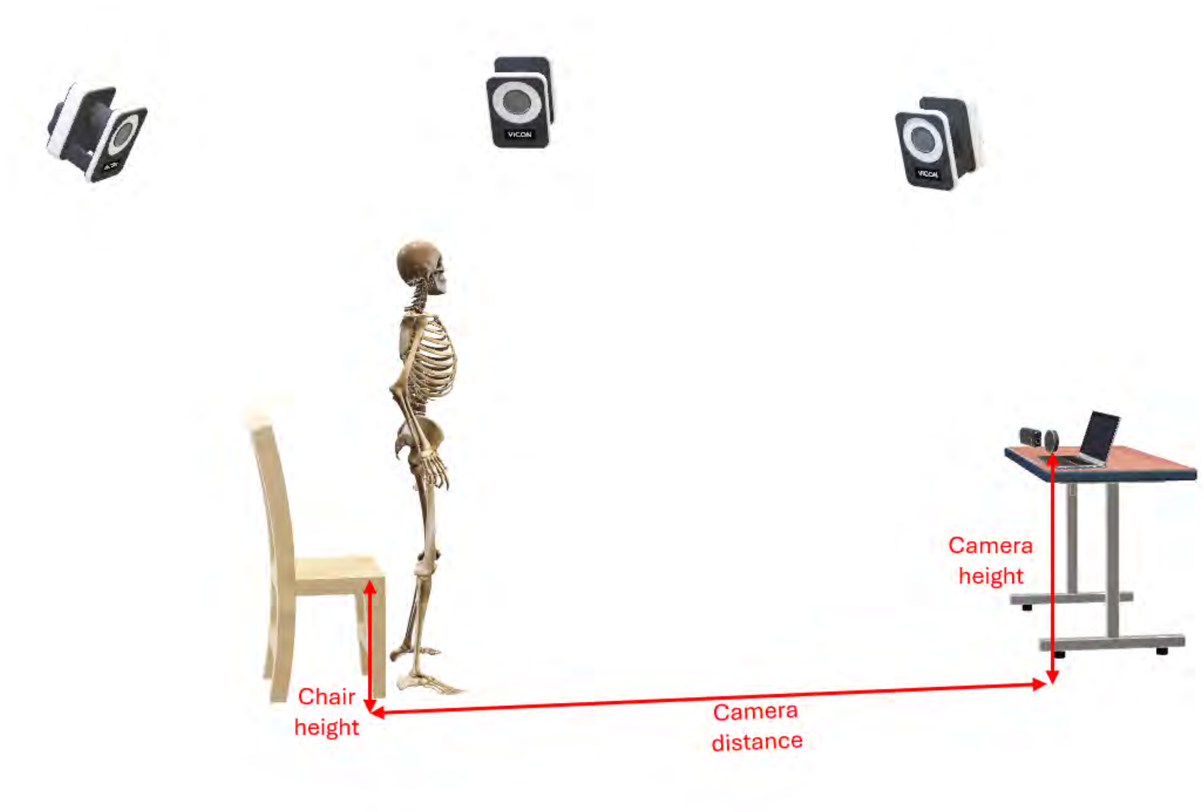
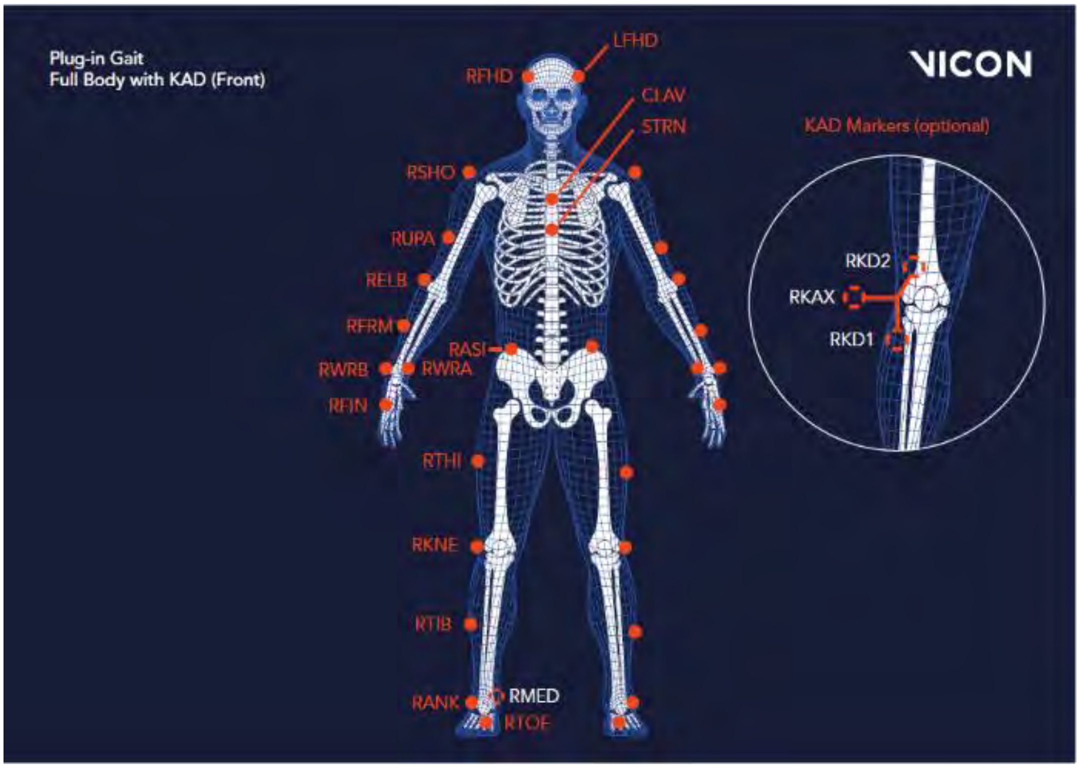


Figure 4.3 Experimental setup showing the position of the two 3D camera systems relative to the chair and a subset of the 9 Vicon cameras surrounding the scene. The chair height is 42 cm, the camera is placed on a 91 cm high surface and the chair was positioned 2.6 m away from the camera.

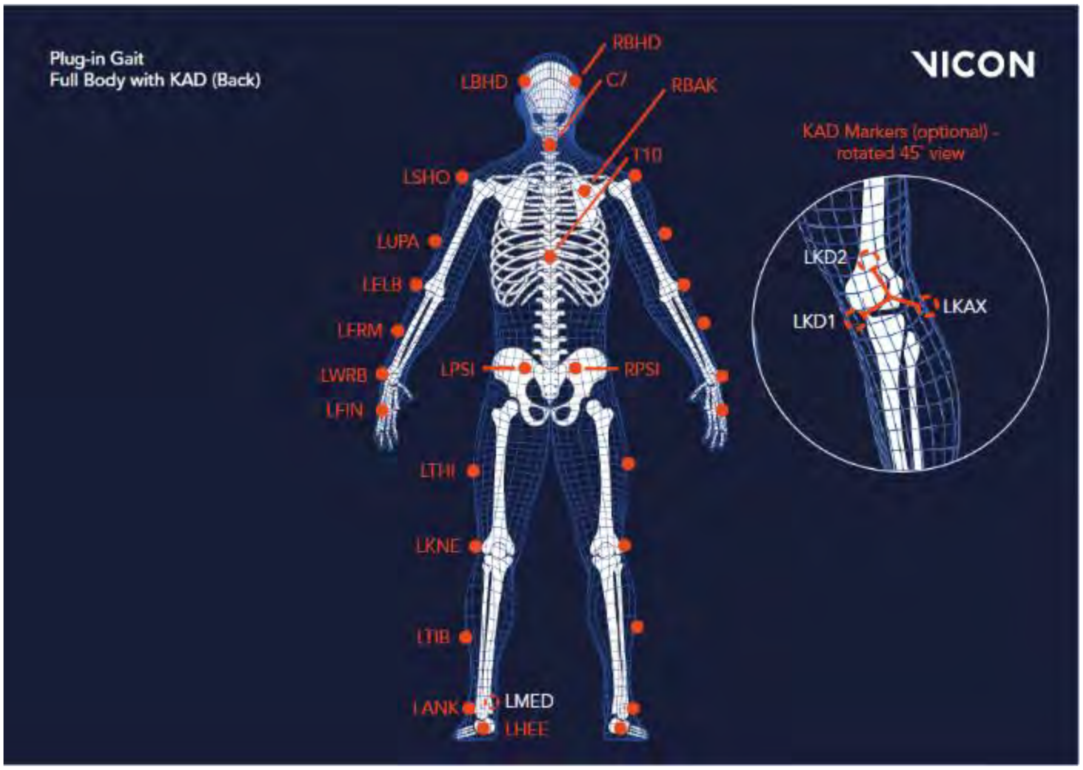
4.3.2 Subject Details

The subject had markers placed on his skin following the marker sets for Plug-in Gait full body modeling from Vicon specified on the official Vicon website (Vicon, 2025a). The IR reflective markers used have a diameter of 14 mm.

Table 4.2 shows the marker placement used. Note that the KAD markers (shown in the detailed views in Figure 4.4) were not used at the knees and that the variation with two posterior superior iliac spine markers on the pelvis was chosen. The subject was wearing high-cut shorts, no shirt, and black shoes.



(a)



(b)

Figure 4.4 Marker sets for Plug-in Gait full body modeling front (a) and back (b) (Vicon, 2025a)

4.4 Experiment

The experiment consists of 11 conditions of three to seven trials (repetitions). The trials for each condition were performed following the same guidelines. Each condition has a different combination of characteristics: rate of StS, body position, body movement or clothing. The conditions are chosen to represent a wide range of potential scenarios. The objective is to find limitations of the system, be it speed, unnatural stance, asymmetrical movements or obstructions. The Table 4.2 describes the conditions used for the experiment. The trials started and ended in a standing position and a metronome was used to guide the StS rate. The whole experiment was conducted on a single day, in less than two hours.

The rate of StS is chosen according to the meta-analysis from Bohannon 2006. The average 5StS time for individuals between the age 60 to 99 is 12.1 seconds, giving a rate of StS of 24.8 repetitions per minute (rpm). Rates of 30, 20, and 18.75 rpm are selected to capture the variability in 5StS time for this population, while ensuring sufficient coverage across different speeds.

The first four blocks are a clean StS without swinging the arms nor using them on armrests, as is usually the case for StS evaluation (Campitelli et al., 2021). The fifth condition is done with exaggerated frontal trunk flexion, using the zero-momentum StS strategy. The sixth condition is with a narrow stance, evaluating if almost overlapping feet hinder their detection. Conditions 07 and 08 test different strategies of foot placement, anterior and posterior respectively (Janssen et al., 2002). Condition 09 is predominantly with a single leg, simulating a weakness or an injury. Finally, conditions 10 and 11 followed the same structure as condition 02 and 03 but with longer, looser fitting shorts, testing the impact of looser clothing, without hiding any markers from the Vicon cameras.

The conditions are generally arranged from easiest to most challenging in three groups. The first group contains standard StS at different speeds with conditions 01 to 04; the second group simulates movements observed in clinics with patients suffering from mobility impairment with conditions 05 to 08; and the third group contains highly asymmetrical movements or challenging clothing with conditions 09 to 11.

Table 4.2 Description of the StS experiment per condition

Condition #	Rate (rpm)	Description
01	30	Standard StS at fast speed without arm assistance and natural foot placement
02	25	Standard StS at average speed without arm assistance and natural foot placement
03	20	Standard StS at slow speed without arm assistance and natural foot placement
04	18.75	Standard StS at very slow speed without arm assistance and natural foot placement
05	25	Slowly with exaggerated trunk flexion throughout
06	20	StS with narrow stance, feet nearly touching to test occlusion sensitivity
07	25	StS with exaggerated forward (anterior) foot placement
08	25	StS with minimal trunk flexion and exaggerated backward (posterior) foot placement
09	25	StS with predominant weight on one foot; simulates unilateral weakness
10	25	Standard StS at average speed with long, loose-fitting shorts; tests effect of clothing
11	20	Standard StS at slow speed with long, loose-fitting shorts; tests effect of clothing

4.5 Motion Capture Software

4.5.1 Nuitrack SDK

The data obtained by the D455 and L515 is a colour feed with depth, called RGB-D. The application selected is Nuitrack SDK, a free off-the-shelf software that can be used with RGB-D cameras and that has been trained to extract the joint centres from scenes containing humans. The version 1.12.11 is used for this experiment.

The application Nuitrack SDK can be run in real time, but this experiment requires to run the pose estimation multiple times on the same recording file to evaluate accuracy. To enable this, the recordings are saved as .BAG files, a data type that can be processed by Nuitrack multiple times at will.

Nuitrack provides the 3D position of joint centres as shown in figure 4.5. The orientation of the joint and the confidence level of the accuracy of the joint data is also provided. This data is acquired at every frame for all visible joints.

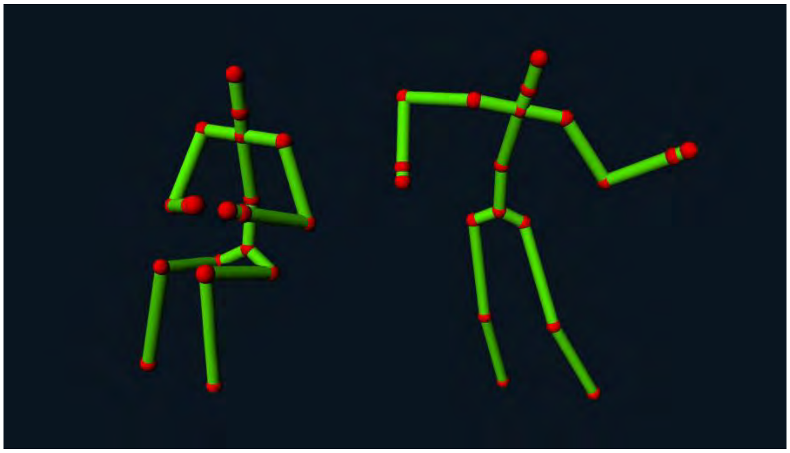


Figure 4.5 Skeleton extraction from Nuitrack. Red spheres are the joints that have an estimated 3D position. (NuitrackSDK, 2024)

4.5.2 Vicon Nexus

Data processing from the Vicon cameras is also done using Vicon Nexus. The Plug-In Gait Model takes as input the data from the static and dynamic marker calibrations as well as anatomical dimensions to process the joint centres from the set of markers. This process is automatic and extracts the joint angles, body joint centre positions and any desired parameter if the input data is without flaws.

Unfortunately, in our experiment, the left hip, femur and knee were sometimes obstructed due to the placement of the chair relative to the cameras. The missing data caused some joint centres positions to be miscalculated in one of the post processing steps to extract joint positions. Joint angles were not affected by this issue and is hence the data that is used to compare the accuracy and precision of the 3D camera systems. The second set of data available for analysis is the 3D position of the markers, the raw data used by Vicon Nexus to calculate the angles.

4.6 Data Processing

4.6.1 Rigid Transformation

The frame of reference of the three systems are proper to them, not only in translation but in rotation as well. The analysis of absolute angles requires these frames of reference to be rotationally aligned. The Vicon system is rotationally aligned with the Z vector being the opposite of the gravitational acceleration vector, but a 90-degree rotation around the Z axis is required to align the

X and Y axes. The desired frame of reference has the X axis pointed anteriorly relative to the subject and the Y axis pointing laterally to the left. The frame of reference is shown in Figure 4.6.

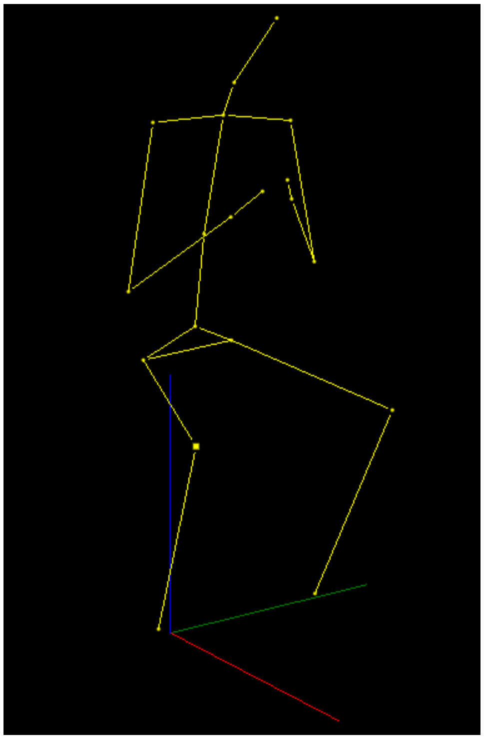


Figure 4.6 Representation of the 3D frame of reference. X in red, Y in green and Z in blue.

In the case of the lidar and stereo cameras, they both contain a 3-axis accelerometer that helps determine the required rotation matrix to align the scene with the global frame of reference. The transformation of every 3D point is processed using the formula below, where P is a position vector, R is a rotation matrix and M is the translation vector.

$$P_{transformed} = \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = \begin{bmatrix} R_{3\times 3} & & & P_{3\times 1} \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_0 \\ y_0 \\ z_0 \\ 1 \end{bmatrix} \tag{4.1}$$

The translation vector is only calculated once to keep the right ankle at the origin, but since the foot placement relative to the chair changes between experimental conditions, the ankle is not always at the origin.

Generally, when evaluating HPE software based on a gold standard, the Mean Per Joint Position Error (MPJPE) is evaluated (Efstratiou & Panagiotis, 2021). This metric requires the 3D transformation to be perfect between the estimated data and the real data. This metric takes the difference between the estimated position and the real position of a joint, for all joints and for all frames, leaving a single scalar value of mean error in position. It is possible to get the MPJPE between the lidar and stereo cameras, but not from the Vicon data, since the position of the joint centres could not be obtained from Vicon Nexus. Differences in position could be compared for specific joints if a marker behaves similarly as a joint centre, but the metric loses its general appeal. For these reasons, the translation vector applied to the estimated positions does not need to be very precise and is mostly used for visual analysis of the three data streams.

4.6.2 Filtering

The data coming from Nuitrack and Vicon pass through a filtering step that removes noise. Filtration is a standard practice when working with 3D data due to the flickering of the 3D position. The filter used is the Savitsky-Golay, used in (Masullo et al., 2019). This filter replaces every point by the intersection of a best fit 3rd degree polynomial of its 12 neighbouring points for our case. This filter is used to keep fast moving features of the StS while removing noise related to the random mispositioning of 3D points.

4.6.3 Interpolation

Interpolation is needed for the comparison of data at a different frame rate. The interpolation transforms discrete data points into a function $f(t) = [x, y, z]$. The SciPy function *Interp1d* is used at this step for each dimension independently through time. The cubic parameter is selected for the interpolation kind, which keeps the first two derivatives of the position continuous at each point.

For the Vicon data, no precise timestamp was available, so the x value for interpolation was based on the framerate of 120 fps from the settings. The first frame being at time 0 and the following 1/120th of a second later. On the other hand, the Nuitrack output from the lidar and stereo cameras have a timestamp precise to the millisecond, which showed that the framerate is not perfectly consistent. Hence, the timestamp was used as the x value for the interpolation step. After acquiring a continuous function for the 3D position, the data is rediscretized at 120 frames per second, to compare all three streams point by point.

4.6.4 Cropping

Data cropping is required to only keep the StS repetitions and none of the walking to and from the computers during the experiment. Cropping is done manually using a graph of the head position in

Z. The Vicon data is cropped definitively in that manner and is only cropped once. On the other hand, the data coming from Nuitrack is cropped twice. A first crop is done to keep the clean StS repetitions by starting in a standing position and ending in a standing position, including a few seconds of buffer at the beginning and at the end of the data. This crop exclusively keeps clean data where all joints are visible and contributes to keeping the files more compact for processing. The second crop happens after the alignment in time, to match the start and end time of the estimated data with the real data. This crop also ensures that the number of points is the same for all three data streams, enabling point-by-point comparison and cross-correlation.

4.6.5 Alignment in Time

The cameras were used on three different computers and without any software synchronization between them, hence aligning the data from the three streams in time is required for any subsequent analysis. The lidar and stereo streams from Nuitrack contain an absolute timestamp that can be used for alignment, but the Vicon data does not, requiring an alternative approach. This approach involves using the cross-correlation coefficient to match the movement curves. The lidar data is shifted in time to test all overlapping possibilities with the Vicon data, calculating the normalized cross-correlation coefficient $\rho(t)$ at every step. The step with the maximum correlation is determined to be the correct alignment and the timestamp is adjusted by the time lag to match the Vicon data.

$$\tau_{lag} = \underset{\tau}{argmax}\, \rho(t) \tag{4.2}$$

The Z component of the head position was used for temporal alignment, as it is experimentally robust to noise and exhibits significant variation over time. This makes it an ideal motion fingerprint for identifying alignment across data streams.

4.7 Joint Angle Extraction

Two types of angles were generated from the joint centre positions obtained by pose estimation: 3-point angles and angles between a line and a plane.

4.7.1 3-Point Angle

A 3-point angle is used to get the angle between two limbs, such as the shank and the femur to get the knee angle. It is not associated with an absolute plane and outputs the smallest angle between two lines. The formula to calculate the angle is below, where · denotes the dot product.

$$\theta = arccos\left(\frac{\overrightarrow{v_1} \cdot \overrightarrow{v_2}}{\|\overrightarrow{v_1}\| \cdot \|\overrightarrow{v_2}\|}\right) \tag{4.3}$$

Following the Vicon standard, the supplementary of the 3-point angle is obtained for the knee with the formula $\theta_{supp} = 180 - \theta$.

4.7.2 Angle Between a Line and a Plane

The angle between a line and a plane is used to describe the orientation of body segments relative to anatomical reference planes, such as the sagittal (XZ) or frontal (YZ) plane. This type of angle is useful for quantifying how inclined a limb is with respect to the global coordinate system.

This angle can be computed using two points p_1 and p_2 to form a line $\vec{v} = p_2 - p_1$. To compute the angle between this line and a plane, the projection of $\vec{v}$ onto the plane is compared with the vertical axis.

For example, to compute the angle relative to the sagittal plane, the formula is shown below. In the equation, arctan2 is the signed version of arctan, that allows using the four quadrants, and v_x and v_z are the X and Z components of the vector $\vec{v}$.

$$\theta = arctan2(v_x, v_z) \tag{4.4}$$

To increase robustness against noise or outliers, especially when multiple points are aligned with the body segment, the line $\vec{v}$ can be fitted using principal component analysis (PCA). This method is used to get the sagittal and frontal trunk angle, since the points in the pelvic region are noisy. The angle is then calculated as above, using the X and Z or Y and Z components of the resulting unit vector, depending on the plane of interest.

4.7.3 Computation of the Trunk Angle

The trunk angle is the only metric calculated with more than three data points. The reason for its special treatment is that most points on the trunk are noisy because of occlusions caused by the arms, making the trunk angle noisier by default. To address this issue, four points aligned with the trunk are selected and used to fit a line using PCA. This approach produces a direction vector that is more robust to noise and outliers than a line defined by only two points, since it averages multiple independent points.

4.8 Data Evaluation

The data acquired from the 3D cameras and Nuitrack must be evaluated and compared to the “gold standard” in multiple ways to assess if it can be used in a clinical setting. This section contains an explanation of the metrics used as well as the evaluation guidelines for accuracy and precision.

4.8.1 Metrics

Metrics are variables obtained from the analysis of the acquired data. They are used clinically since they are more physiologically understandable than pure 3D positions and reflect the task at hand. The metrics selected are values that cannot trivially be computed using a simple 2D video of StS. For example, 5StS is a timed metric that can easily be timed using such a video and that has little to no uncertainty related to the lack of depth data.

The first group of metrics computed from the 3D data contains 10 angles and are shown in Figure 4.7. The shank, knee, femur and hip angles are obtained for the left and right side of the body, with the shank and femur angles being absolute angles projected to the sagittal plane. Finally, the trunk angle is obtained in the sagittal and frontal planes. All the angles shown in Figure 4.7 tend to 0° when reaching a standing position and all angles increase when sitting, except for the femur angle, which reaches a negative value.

The 3-point angle method in section 4.7.1 is used for two metrics: the knee angle uses the ankle, knee and hip joints; and the hip angle uses the knee, hip and torso. The angle between two points and a plane method in section 4.7.1 is used for two metrics: the shank uses the ankle and the knee; and the femur uses the knee and the hip. The trunk angle is slightly more complex and is explained in section 4.7.3

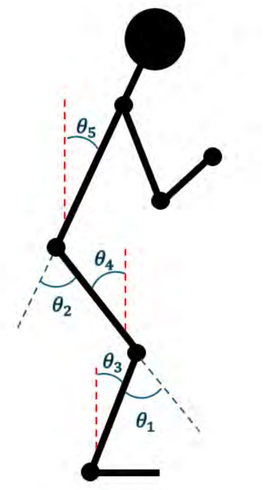


Figure 4.7 Angles calculated in the sagittal plane. The blue dotted lines represent lines that follow the direction of the previous body part, and the red dotted lines are vertical relative to the ground. $\theta_1$is the knee angle, θ_2 is the hips angle, θ_3 is the sagittal shank angle, θ_4 is the sagittal femur angle and θ_5 is the sagittal trunk angle.

4.8.2 Computation of Velocity and Acceleration Metrics

The second and third group of metrics computed are velocity and acceleration metrics. Velocities are obtained from metric angles from the first group of metrics. Two additional velocities are computed from the upwards (Z) and forward (X) component of the head position. Accelerations are obtained from all velocity metrics.

A short window of 20 frames is applied to smooth the position data before computing velocities, as direct differentiation amplifies noise. At 120 fps, this window spans 1/6th of a second, balancing noise reduction with temporal resolution. Then the gradient of the points is computed giving the velocity at that timeframe.

Computation of acceleration is done in an identical manner, computing the gradient of the velocities obtained at the previous step.

4.8.3 Precision Evaluation Guidelines

The pose estimation performed by Nuitrack is nondeterministic; running the software multiple times on the same input file yields slightly different 3D joint positions. To quantify the system's precision, Nuitrack was executed 30 times per recording. According to the central limit theorem, a sample size of 25 to 30 is typically sufficient to approximate the sampling distribution of the mean as normal, even when the underlying distribution is unknown (Hogg et al., 2020).

However, due to outliers and asymmetrical errors, relying solely on the mean and the standard deviation can hide the true tendencies of the data. Instead, percentile-based statistics provide a more robust overview. After the standard preprocessing detailed above, the percentile-based statistics are calculated for each data point across the 30 files. These statistics are the 2.5^{th}, 50^{th} and 97.5^{th} percentile, representing the middle of the distribution and a confidence interval of two sigmas.

The metrics (such as the knee or trunk angles) used for comparison with the ground truth values could be calculated directly from the median and percentile values across the 30 software iterations. However, since the data points within each run are not assumed to be independent, the metrics are computed individually for each run. The metric percentiles are then calculated from the distribution of these per-run metric values.

Precision between software iterations is evaluated with CCC, shown in section 4.8.4, to evaluate the similarity in shape and values. CCC only compares pairs of values, so a correlation matrix is plotted as well as violin plots for various metrics. Based on the 95% confidence interval of the correlation coefficient estimate, values below 0.5 indicate poor reliability; values between 0.5 and 0.75 indicate moderate reliability; values between 0.75 and 0.90 indicate good reliability; and values above 0.90 indicate excellent reliability (Koo & Li, 2016).

4.8.4 Accuracy Evaluation Guidelines

Accuracy is a measurement of how close the data is to the true value. In the case of motion capture for rehabilitation, the acceptable error is generally established to be 5° for joint angles (Feldhege et al., 2015; McGinley et al., 2009; Riazati et al., 2022). The RMSE over a whole sequence, the value needed is usually around 5° (Kiatkulanusorn et al., 2023). During rapid multi-joint motions, the acceptable peak error seems to increase to around 10° if the error spans only a few frames (Hullfish et al., 2019). In general, an error of less than 3° is to be expected while an error of more than 5° is cause for concern (Riazati et al., 2022).

Accuracy for angles will be evaluated using the residuals, Mean Joint Angle Error (MJAE), RMSE, Pearson’s r, CCC and Percentage of Correctly Estimated Angles at $\epsilon°$ (PCEA-$\epsilon°$).

The MJAE is calculated with the formula below, where θ_i and $\hat{\theta}_i$ are the real angles and the estimated values respectively for a single joint and N is the number of frames.

$$MJAE = \frac{\sum_{i=0}^{N}\left(\theta_i - \hat{\theta}_i\right)}{N} \tag{4.5}$$

The RMSE is calculated in a similar fashion but penalizes more heavily high variation of errors than MJAE.

$$RMSE = \sqrt{\frac{\sum_{i=0}^{N}\left(\theta_i - \hat{\theta}_i\right)^2}{N}} \tag{4.6}$$

Pearson’s r is a correlation coefficient that measures similarity in shape. It shows how linearly correlated two continuous variables are without any regard for a systematic bias in the data. The Pearson r is calculated with the formula below. The variables x and y represent the estimated and real values, the subscript i represents the data point, and μ is the mean of the vector.

$$r = \frac{\sum_{i=0}^{n}(x_i - \mu_x)\left(y_i - \mu_y\right)}{\sqrt{\sum_{i=0}^{n}(x_i - \mu_x)^2} \cdot \sqrt{\sum_{i=0}^{n}(y_i - \mu_y)^2}} \tag{4.7}$$

The Concordance Correlation Coefficient (CCC) is a correlation coefficient measuring the similarity in shape and the similarity in values of two continuous variables. It is calculated using the formula below where, r is the Pearson’s r as calculated above, σ is the standard deviation of the vector and μ is its mean.

$$CCC = \frac{2r\sigma_x\sigma_y}{\sigma_x^2 + \sigma_y^2 + \left(\mu_x - \mu_y\right)^2} \tag{4.8}$$

The PCEA-ϵ° is a percentage value of how many estimated are within ϵ° of the real value. The indicator function $\mathbb{1}$, equals 1 if the condition is true and 0 otherwise (Efstratiou & Panagiotis, 2021).

$$PCEA = \frac{1}{n}\sum_{i=0}^{n} \mathbb{1}\left(\left|\theta_i - \hat{\theta}_i\right| < \epsilon\right) \tag{4.9}$$

PCEA is also plotted to give the Cumulative Accuracy Curve where PCEA is plotted as a function of ϵ.

The last plot type is the residual plot, where the real value lies on the x-axis and the residual, the difference between the real and the estimated value, lies on the y-axis. This plot identifies the movement regions where the estimation errors are most pronounced.

Accuracy for the velocity and acceleration will be computed using the peak residual. The peak values are computed using *find.peaks* from *scipy.signal.* A minimum distance between peaks is enforced to keep a single peak per StS repetition, maximum and minimum. The distance is set to 85% of the period of the StS motion. 85% was chosen experimentally to balance a slight variation in StS timing. The residual per frame is calculated with:

$$Residual_{frame} = x_{frame} - \hat{x}_{frame} \tag{4.10}$$

The relative residual is calculated with:

$$Residual_{relative} = \frac{x_{frame} - \hat{x}_{frame}}{\hat{x}_{frame}} \tag{4.11}$$

4.9 Summary and Processing Pipeline

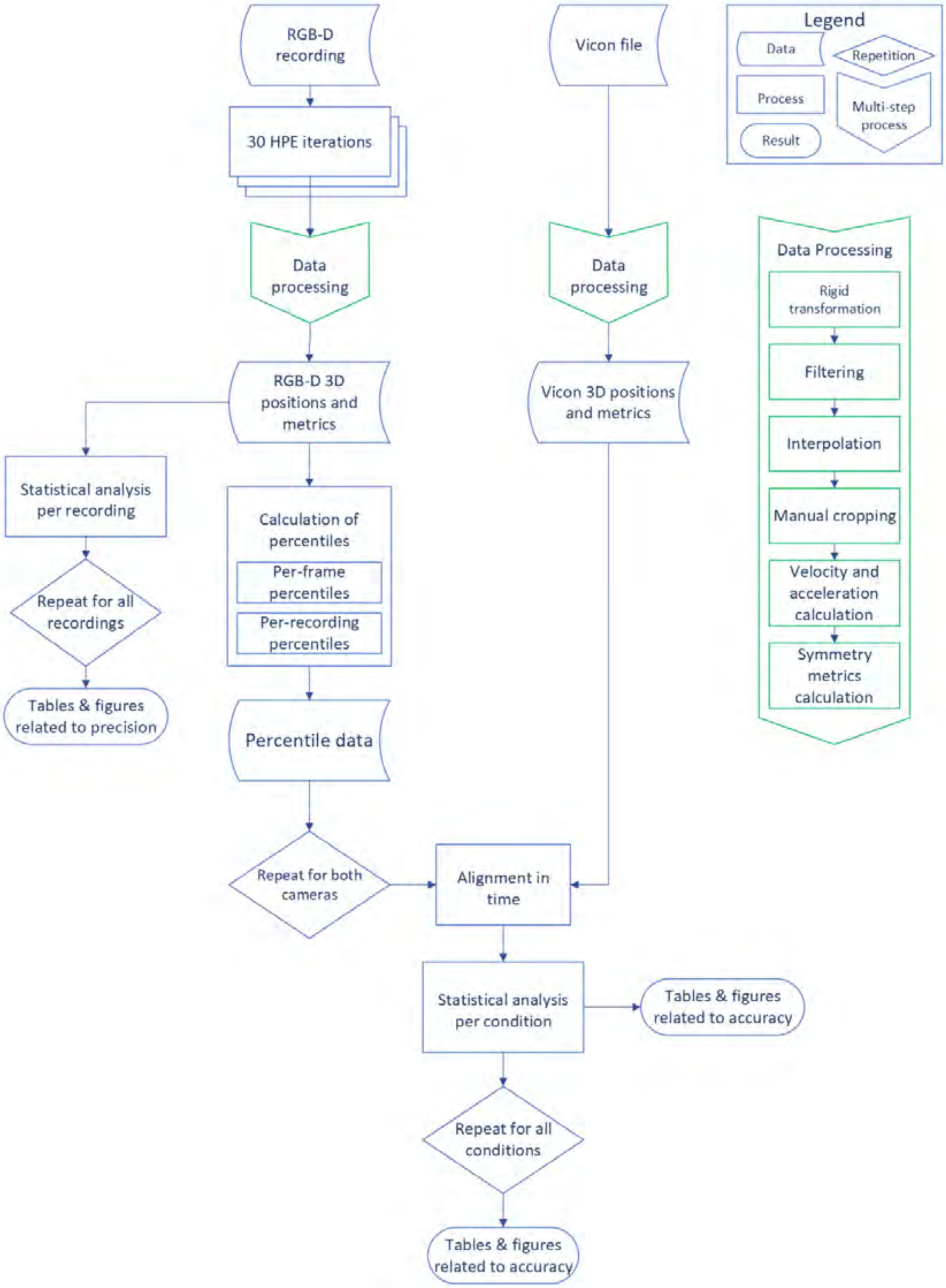


Figure 4.8 Bloc diagram summarizing sections 4.6 to 4.8.

Figure 4.8 illustrates the full data analysis pipeline, from raw input files, through intermediate data structures and per-run/condition-level summaries, to the generation of final evaluation metrics and plots used in Chapter 5.

4.9.1 Precision Evaluation Summary

The evaluation of precision begins with an RGB-D recording from one of the 3D cameras. There is one recording per condition per camera. The recording is run 30 times through the HPE software, and the file containing the joint centre at each frame is saved for further analysis.

The angles are calculated from these 3D positions and summary metrics (such as the symmetry between a right and left joint) is computed from the angles. Having found all the values from a single run, the remaining 29 can be processed in the same way.

Now that the data for all software iterations of a given recording is computed, correlations between the software iterations can be found, giving results for the precision evaluation section. This data is now used in two different ways, first to find the percentile statistics and second to find tendencies across multiple recordings.

The percentile statistics are computed for every 3D point, metric and summary metric, reducing the amount of data being carried while keeping valuable information about the spread of the data for this single recording. This percentile file is later used as an input for the accuracy evaluation section.

The second way that the data from multiple software iterations is used is by merging it with the data obtained from all the other recordings. There are 11 conditions and two 3D cameras which gives 660 total software iterations that can be compared and correlated. With this data, general precision trends about the cameras, the conditions and the metrics computed can be computed.

4.9.2 Accuracy Evaluation Summary

The percentile file computed above is one of the two inputs needed for this phase. The second input is the Vicon Data. The raw output from Vicon Nexus is converted into a similar format as the percentile file for a seamless comparison.

Each condition has two percentile files, L515 and D455, and a single Vicon file. The three files for are cropped to the exact same length and aligned to find the error at every point. At this stage, a single condition is evaluated in terms of how closely the estimation matches the ground truth.

The final step is to merge the data from all 11 conditions to find accuracy tendencies about the cameras, the conditions and the metrics computed.

CHAPTER 5 RESULTS

The results are divided in three sections. Chapter 5.1 begins with preliminary observations related to specific problems of the Human Pose Estimation (HPE) data, of the ground truth, or failures related to the methodology. Addressing these concerns first provides context for the remainder of the analysis and lets the following sections focus on other key findings. Chapter 5.2 follows with observations related to the precision of the HPE. This section presents the repeatability of the estimations for the L515 and D455 when the same 3D recording is analyzed multiple times. Chapter 5.3 concludes with observations related to the accuracy of the HPE, by comparing the estimations to the Vicon ground truth.

5.1 Preliminary Observations

5.1.1 Ankle Position and Shank Angle

First, the position of the ankle using the 3D cameras was often lost when in the upright position. For the frames where the ankle is lost, as shown in Figure 5.1, the angles relying on the position are skewed, showing a sudden disconnect in the angle data.

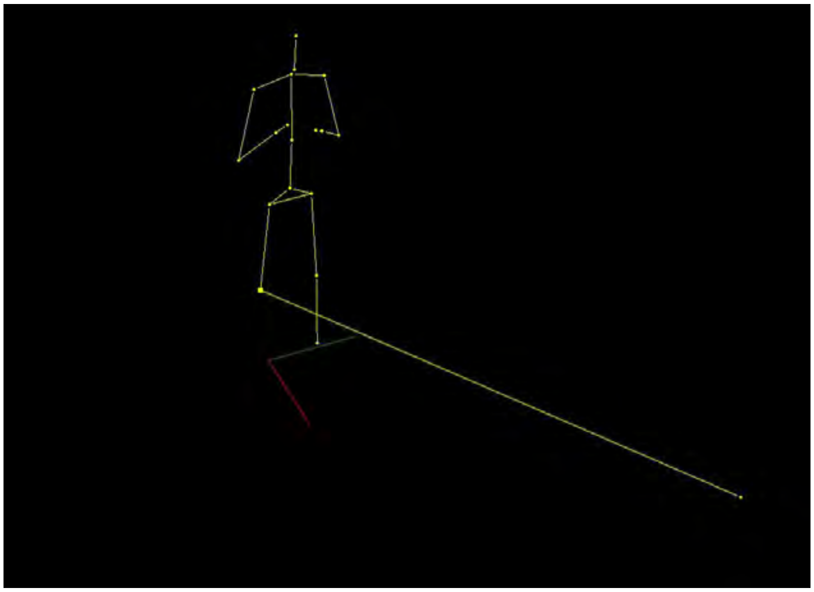


Figure 5.1 Loss of the right ankle position for the condition 10 with the D455. The position of the ankle is assumed to be at the origin of the scene when Nuitrack loses its position.

Figure 5.2 and Figure 5.3 show an easy condition (01) and a challenging condition (10), filmed with the L515 and D455 respectively, where this type of error can be seen in the latter but not the former. The Table 5.1 shows the occurrence of this type of error across the conditions classified by camera and body side. The count is highest for the right side using the D455 followed by the left side using the L515.

Table 5.1 Count of the loss of ankle position per recording, split by laterality. Loss of ankle position on one side affects the shank angle and the knee angle, since they both rely on the ankle position.

Description	Camera	L515		D455	
	Body side	Right	Left	Right	Left
Standard StS at fast speed without arm assistance and natural foot placement	Condition 1	0	0	1	0
Standard StS at average speed without arm assistance and natural foot placement	Condition 2	0	1	5	0
Standard StS at slow speed without arm assistance and natural foot placement	Condition 3	0	1	7	0
Standard StS at very slow speed without arm assistance and natural foot placement	Condition 4	0	1	4	0
Slowly with exaggerated trunk flexion throughout	Condition 5	0	0	1	0
StS with narrow stance, feet nearly touching to test occlusion sensitivity	Condition 6	0	0	0	0
StS with exaggerated forward (anterior) foot placement	Condition 7	0	0	0	0
StS with minimal trunk flexion and exaggerated backward (posterior) foot placement	Condition 8	0	4	1	0
StS with predominant weight on one foot; simulates unilateral weakness	Condition 9	0	0	1	0
Standard StS at average speed with long, loose-fitting shorts; tests effect of clothing	Condition 10	0	0	4	4
Standard StS at slow speed with long, loose-fitting shorts; tests effect of clothing	Condition 11	4	4	10	0

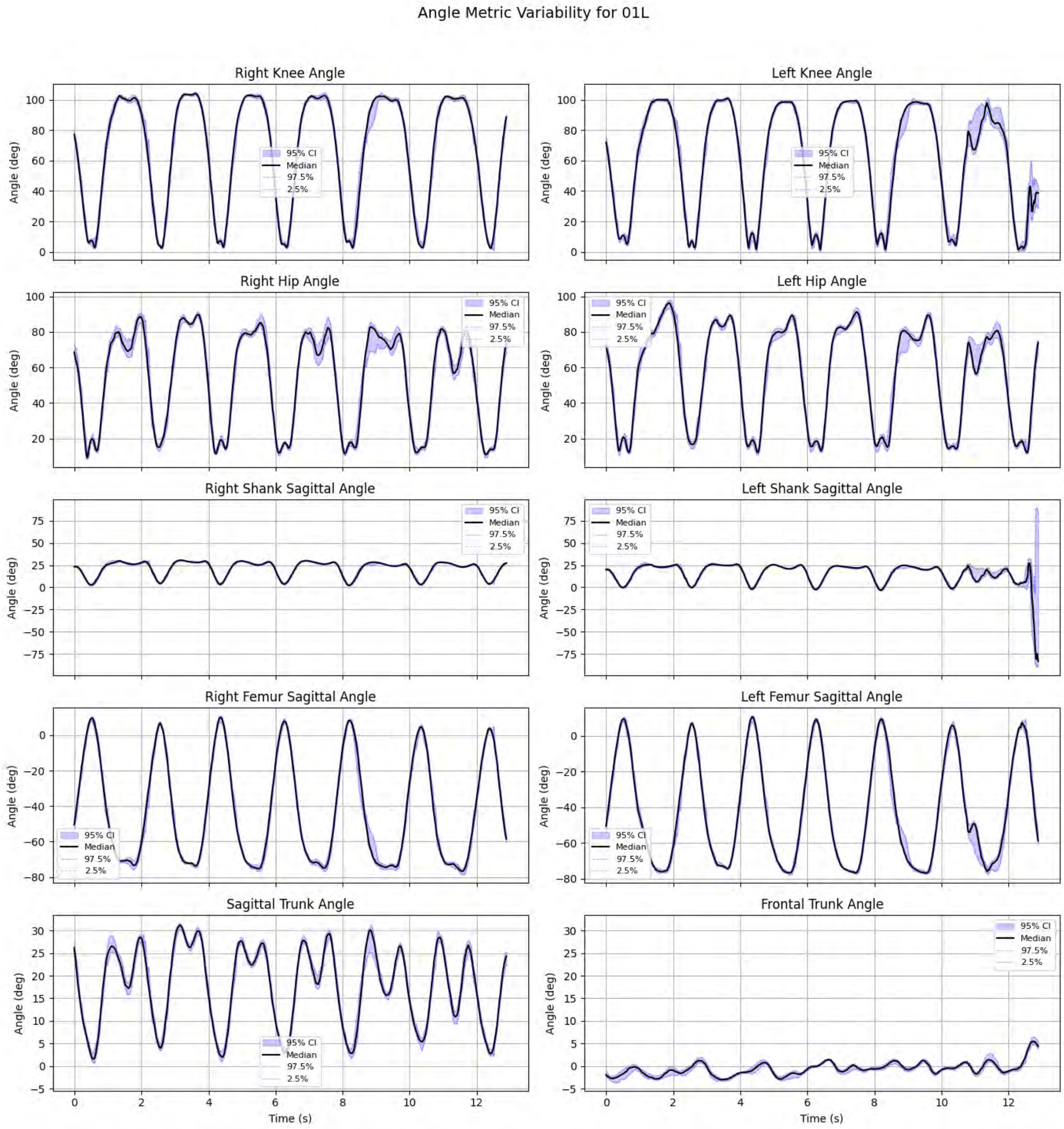


Figure 5.2 Results of angle measurements for condition 01 with the L515, showing generally clean shank angle data except in the last second on the right side. All other angles are also continuous, without outliers. The black lines show the median value of the 30 software iterations and the blue fill lines show a 95% confidence interval using the percentiles 2.5 and 97.5.

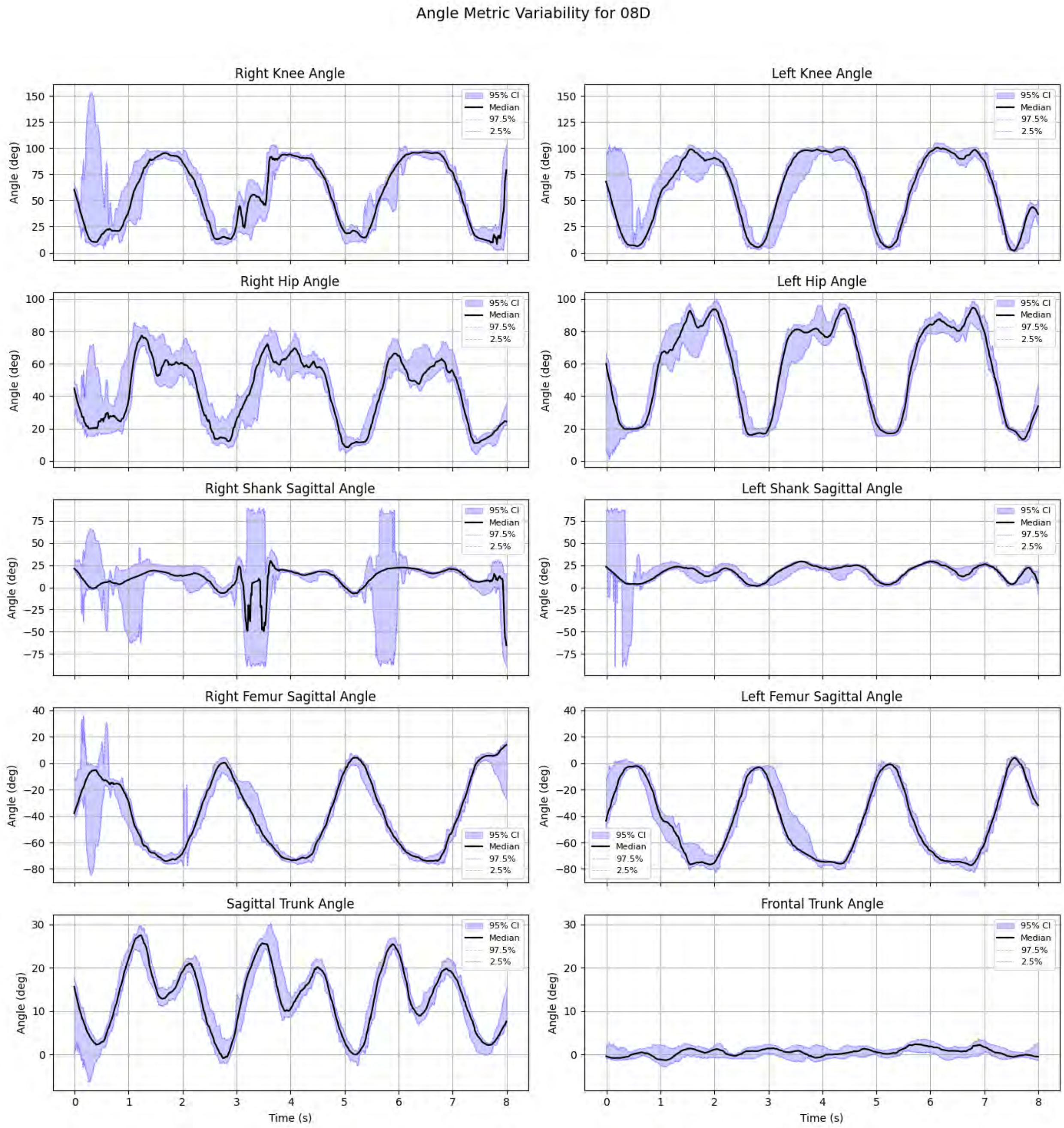
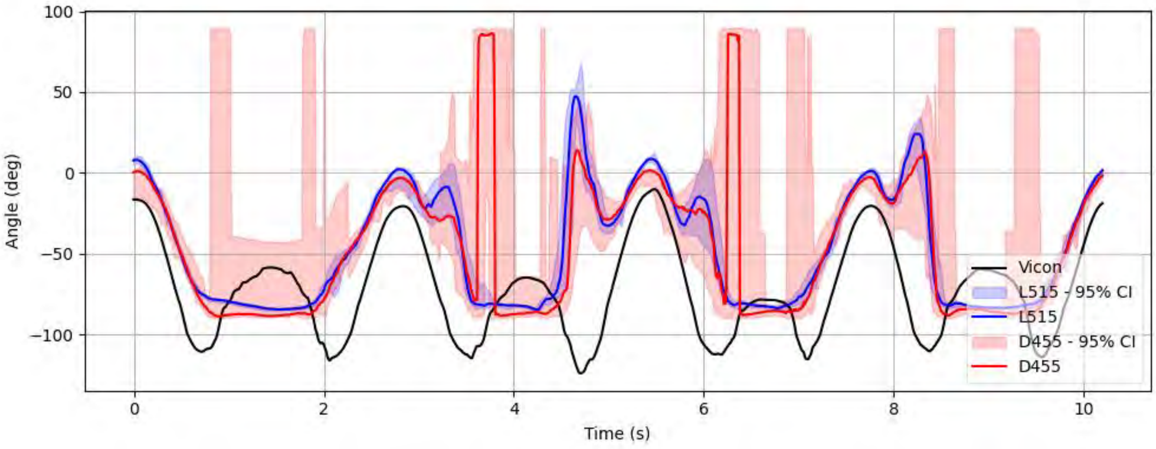


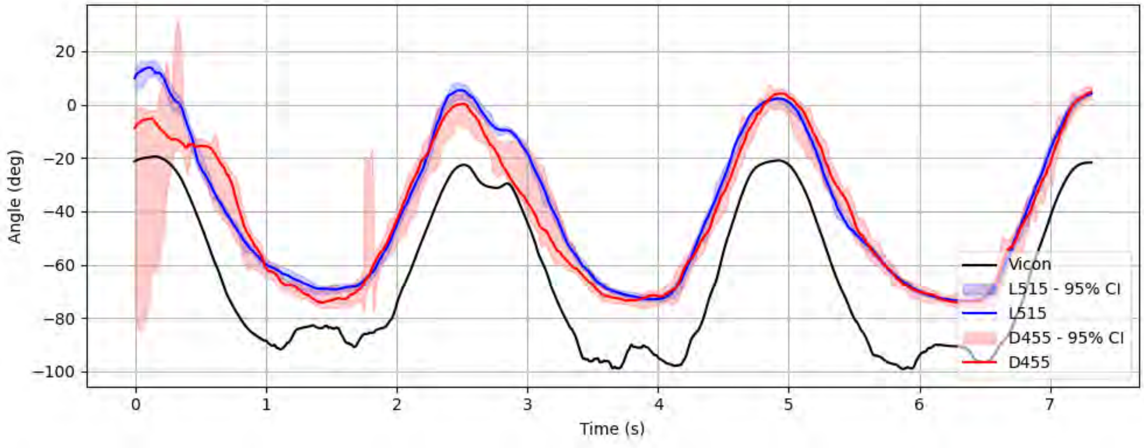
Figure 5.3 Results of angle measurements for condition 10 with the D455 (10D), showing problematic shank angles on the right and left sides. The knee angle becomes problematic as well when the shank angle is wrong because the knee angle is calculated with the femur and shank segment. These errors do not impact any other angle, as shown in this figure. See Figure 5.2 for details on the figure.

5.1.2 Sagittal Femur Angles

The sagittal femur angle for the Vicon data follows an unusual trajectory between seat touch-down and seat lift-off, when the femur angle reaches its minimum value in Figure 5.4. In between these two points, the angle should remain approximately constant since the knee and hip joints have a static height when the subject is seated down during phase one of the StS motion. The femur angle measured with Vicon instead increases and decreases by around 5 to 50 degrees depending on the condition. At the same time, the femur angle remains generally constant for the L515 and D455 data.



a)

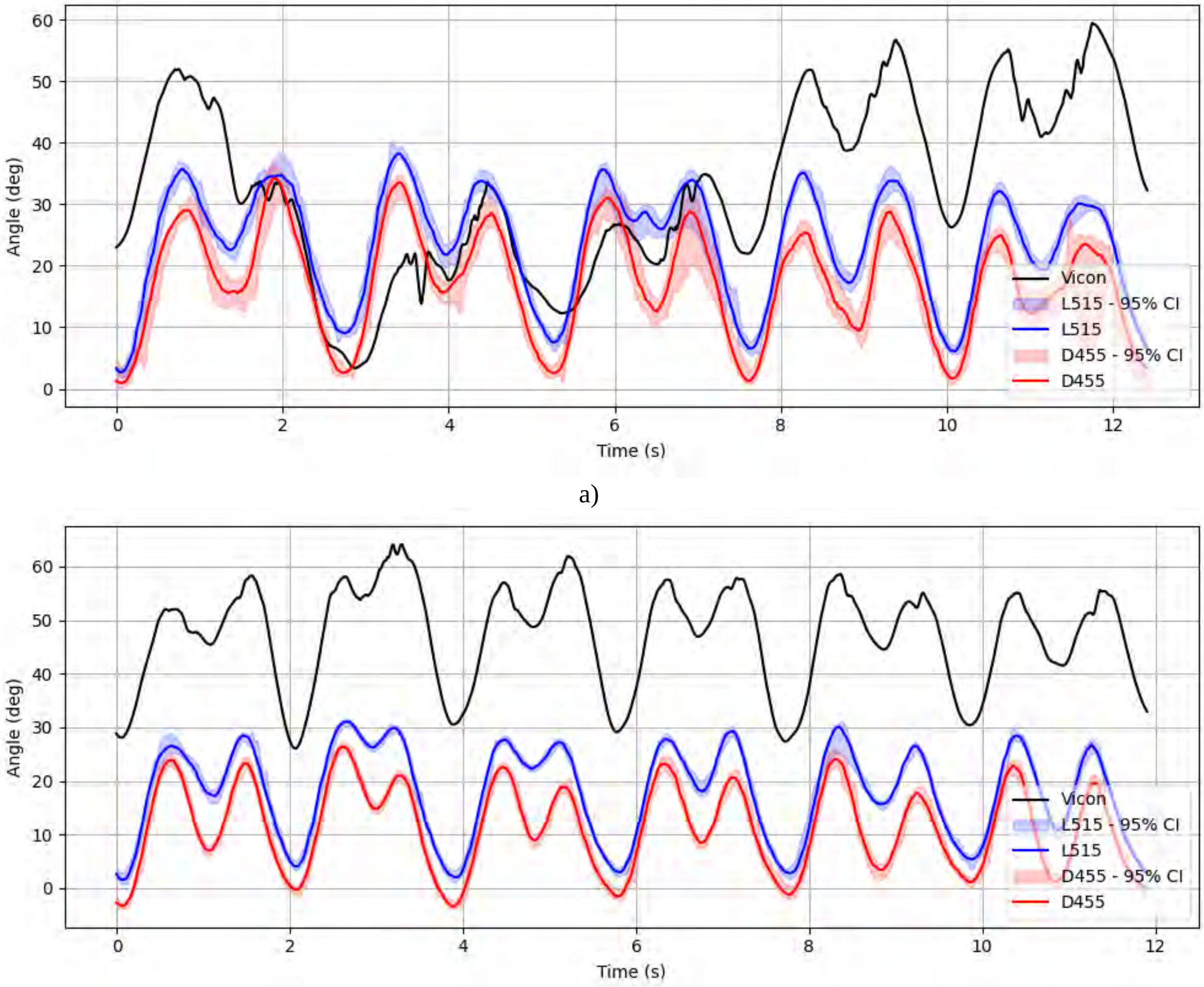


b)

Figure 5.4 Right sagittal femur angle comparison with Vicon, the L515 and the D455 data for conditions a) 07 and b) 08. Condition 07 shows a large peak when the angle reaches the minimum but condition 08 shows a very small peak. The main difference between these conditions is the foot placement and the need for a large sagittal trunk movement.

5.1.3 Problematic Sagittal Trunk Angle Ground Truth for Condition 06

A single condition showed an issue with the sagittal trunk angle. Indeed, the average trunk angle measure is well above the trunk angle measure by Nuitrack by about 30°. This systematic error appears in all StS repetitions for all conditions except two repetitions in condition 06, where the trajectory drops by around 30°. Condition 06 and a reference figure are shown in Figure 5.5.



a)

b)

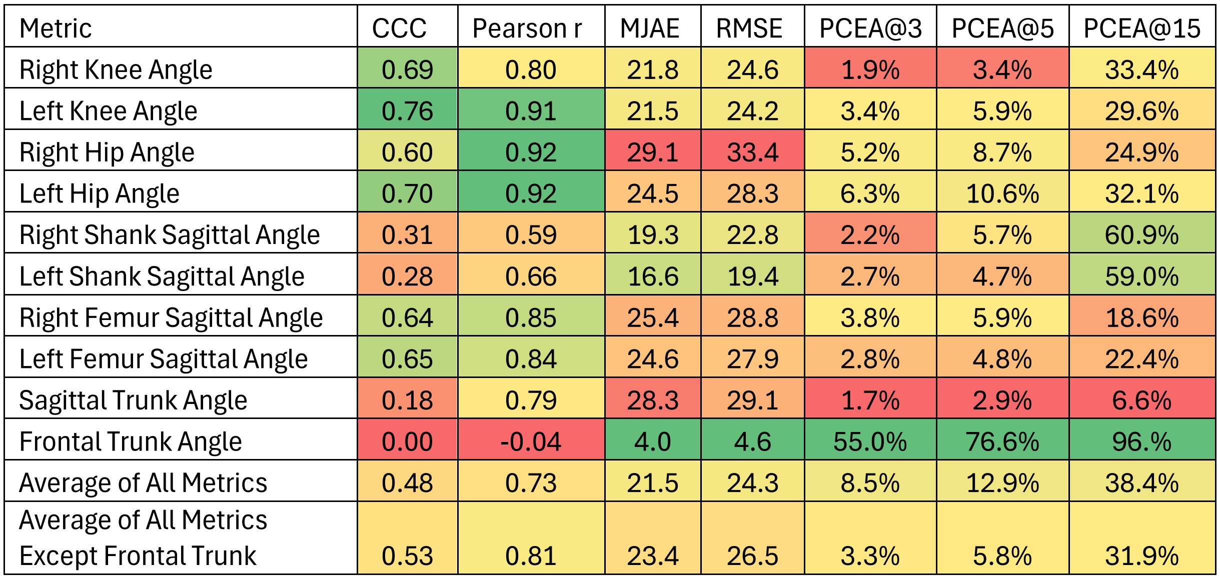
Figure 5.5 Examples of the sagittal trunk angle from all three sources. a) shows the problematic condition 06 where the offset disappears for two StS repetitions and b) shows the systematic 30° offset in condition 01 present in other conditions as well.

5.1.4 Frontal Trunk Angles

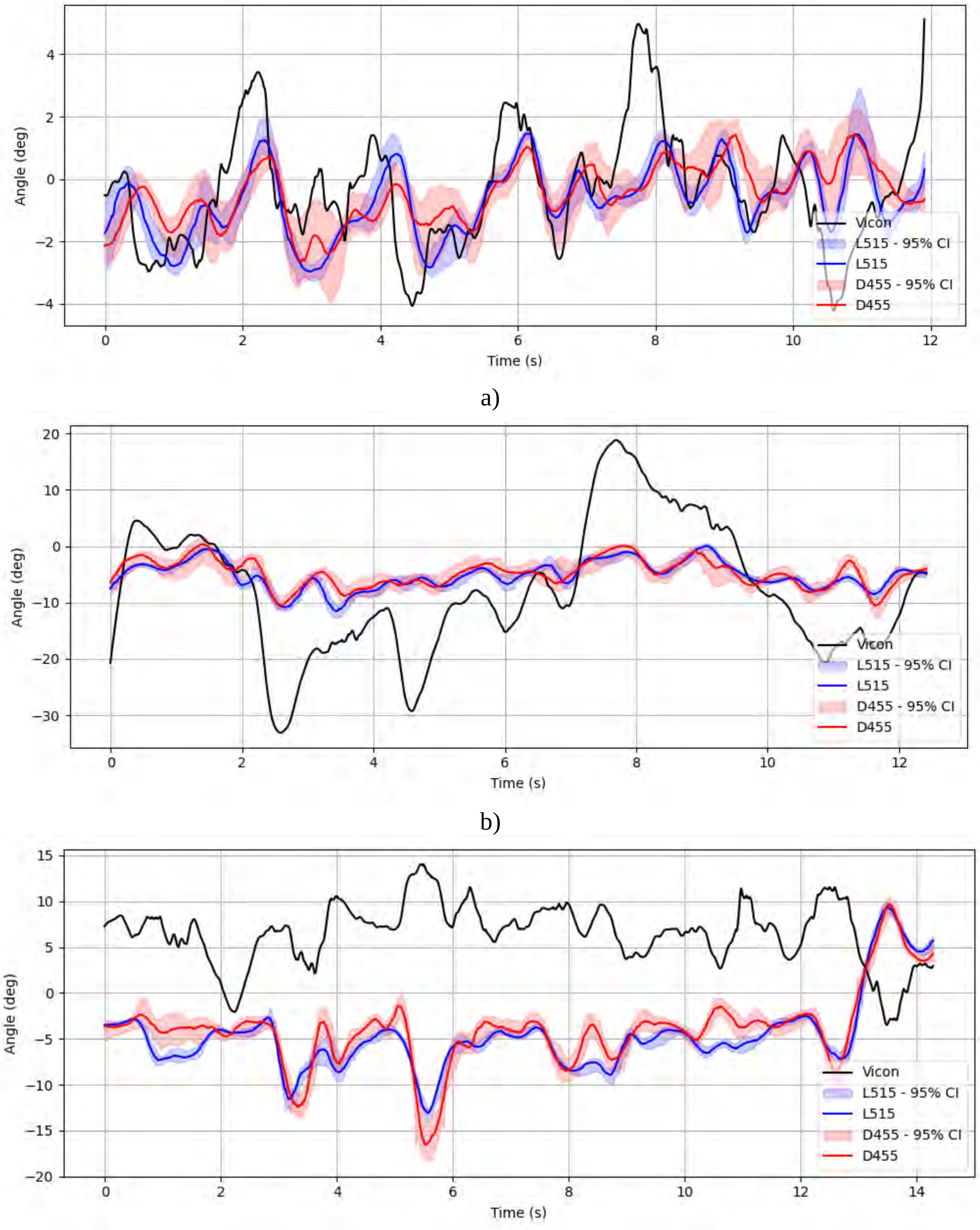
The frontal trunk angle has been computed for every run, but the StS movements do not happen in the frontal plane, meaning that the signal-to-noise ratio is very small. As shown in Table 5.2, the

MJAE is the smallest, but the accuracy Concordance Correlation Coefficient (CCC) is 0, meaning that there is no relationship between the data and the ground truth. This contradiction shows that the experiment was not well suited to evaluating the frontal trunk angle. For this reason, the frontal trunk angle will not be included in the results below to keep the averages representative of the trends obtained with the properly evaluated metrics. An overview of the results related to the frontal trunk angle is presented in Figure 5.6.

Table 5.2 Statistical results per metric averaged over all conditions and both cameras. The frontal trunk angle has the best MJAE, RMSE and PCEA at every ϵ, but the CCC and Pearson r are the worst of all metrics. The table is colour coded continuously from green to red with green being the best relative values per metric and red being the worst.



Metric	CCC	Pearson r	MJAE	RMSE	PCEA@3	PCEA@5	PCEA@15
Right Knee Angle	0.69	0.80	21.8	24.6	1.9%	3.4%	33.4%
Left Knee Angle	0.76	0.91	21.5	24.2	3.4%	5.9%	29.6%
Right Hip Angle	0.60	0.92	29.1	33.4	5.2%	8.7%	24.9%
Left Hip Angle	0.70	0.92	24.5	28.3	6.3%	10.6%	32.1%
Right Shank Sagittal Angle	0.31	0.59	19.3	22.8	2.2%	5.7%	60.9%
Left Shank Sagittal Angle	0.28	0.66	16.6	19.4	2.7%	4.7%	59.0%
Right Femur Sagittal Angle	0.64	0.85	25.4	28.8	3.8%	5.9%	18.6%
Left Femur Sagittal Angle	0.65	0.84	24.6	27.9	2.8%	4.8%	22.4%
Sagittal Trunk Angle	0.18	0.79	28.3	29.1	1.7%	2.9%	6.6%
Frontal Trunk Angle	0.00	-0.04	4.0	4.6	55.0%	76.6%	96.%
Average of All Metrics	0.48	0.73	21.5	24.3	8.5%	12.9%	38.4%
Average of All Metrics Except Frontal Trunk	0.53	0.81	23.4	26.5	3.3%	5.8%	31.9%



a)

b)

c)

Figure 5.6 Examples of the frontal trunk angle from all three sources: a) shows condition 01 which has the best CCCs (L515: 0.39 and D455: 0.19), b) shows a failure mode in condition 06 where the ground truth diverges from the vertical even though the movement followed the frontal plane and c) shows a second failure mode in condition 09 where the trajectories seem to be inverted as shown by negative and lowest CCCs (L515: -0.13 and D455: -0.11).

5.2 Observations on the Precision of Measurements?

The precision results are related to the repeatability of the HPE. This section presents figures and tables reporting statistical measures of agreement between multiple HPE software iterations for the same recording.

5.2.1 Precision of Angle Metrics

Table 5.3 shows the CCC values averaged over the 30 software iterations for every recording, evaluating the agreement between multiple software iterations of the HPE for a single recording. The CCC values vary between 0.35 and 1.00 for individual metrics, a large spread ranging from poor to excellent. The average per file varies between a good value of 0.79 and an excellent value of 0.99 and the average per metric for all files varies between a similar 0.77 and 0.96. When comparing both cameras, the average CCC for all metrics is 0.94 for the L515 and 0.89 for the D455. When comparing individual metrics, the lowest CCC values are for the shank angles and the frontal trunk angles with values of 0.77 to 0.87, while all other angles have an average above 0.93.

The conditions are ordered in three general groups from easiest to most challenging as explained in section 4.4 and the CCC tends to get lower for conditions that are more challenging. In Table 5.3, a lot of green is observed for the first group of conditions one to four, while almost no green is obtained for conditions 10 and 11, with central conditions having a mix of green and yellow.

The density functions of the CCC results are shown in Figure 5.7. Where Table 5.3 presented the mean CCC, this figure presents the density function from the CCC values found in the 30x30 correlation matrix of each recording. The spread for most metrics is much smaller for the L515 than with the D455, except for the left shank angle, as we know from Table 5.1 showing that the left ankle is problematic for the L515. The D455 data generally has a very long tail towards low CCC values, showing that a single or a few software iterations of HPE for a single or few conditions had a very poor correlation with the overwhelming majority of runs.

Table 5.3 Average CCC values per angle per recording for all 30 software iterations for the L515 (L) and the D455 (D). The results are colour coded from red to green continuously according to their values. (<0.5 = poor, 0.5-0.74 = moderate, 0.75-0.89 = good, >0.9 = excellent)

File	Right Knee	Left Knee	Right Hip	Left Hip	Right Shank Sagittal	Left Shank Sagittal	Right Femur Sagittal	Left Femur Sagittal	Sagittal Trunk	Frontal Trunk	All Metrics
01L	1.00	0.99	0.99	0.99	0.99	0.98	1.00	1.00	0.99	0.92	0.98
01D	0.98	0.99	0.99	0.99	0.70	0.98	1.00	1.00	0.99	0.85	0.95
02L	0.99	0.98	0.98	0.99	0.99	0.63	0.99	0.99	0.98	0.97	0.95
02D	0.88	0.99	0.97	0.97	0.73	0.98	0.99	0.99	0.97	0.88	0.94
03L	1.00	0.99	0.99	0.99	0.99	0.47	1.00	1.00	0.99	0.98	0.94
03D	0.79	1.00	0.96	0.99	0.63	0.99	0.92	1.00	0.98	0.84	0.91
04L	1.00	0.98	1.00	0.99	0.99	0.67	1.00	0.99	0.99	0.95	0.96
04D	0.90	0.99	0.97	0.98	0.38	0.97	0.99	0.99	0.96	0.88	0.90
05L	0.98	0.98	0.92	0.94	0.77	0.79	0.93	0.84	0.98	0.81	0.89
05D	0.98	0.97	0.92	0.94	0.75	0.97	0.93	0.83	0.98	0.81	0.91
06L	0.99	0.99	0.98	0.98	0.98	0.98	0.99	0.99	0.98	0.97	0.98
06D	0.99	0.94	0.96	0.96	0.97	0.91	0.99	0.94	0.94	0.86	0.95
07L	0.99	0.90	0.95	0.94	0.99	0.55	0.96	0.98	0.98	0.88	0.91
07D	0.92	0.91	0.87	0.86	0.92	0.91	0.52	0.58	0.92	0.76	0.82
08L	1.00	1.00	1.00	1.00	0.99	0.99	1.00	1.00	1.00	0.94	0.99
08D	0.88	0.97	0.92	0.97	0.31	0.65	0.93	0.98	0.94	0.68	0.82
09L	1.00	1.00	0.99	0.99	1.00	0.99	1.00	1.00	0.99	0.99	0.99
09D	0.99	0.96	0.96	0.96	0.77	0.95	0.99	0.96	0.98	0.97	0.95
10L	0.85	0.87	0.94	0.96	0.59	0.63	0.96	0.94	0.99	0.93	0.87
10D	0.78	0.78	0.93	0.96	0.46	0.49	0.99	0.96	0.96	0.60	0.79
11L	0.85	0.85	0.93	0.89	0.63	0.62	0.94	0.94	0.91	0.80	0.84
11D	0.78	1.00	0.92	0.98	0.35	0.98	0.80	1.00	0.98	0.84	0.86
All L	0.97	0.96	0.97	0.97	0.90	0.75	0.98	0.97	0.98	0.92	0.94
All D	0.90	0.95	0.94	0.96	0.63	0.89	0.91	0.93	0.96	0.82	0.89
All Files	0.93	0.96	0.96	0.96	0.77	0.82	0.95	0.95	0.97	0.87	0.91

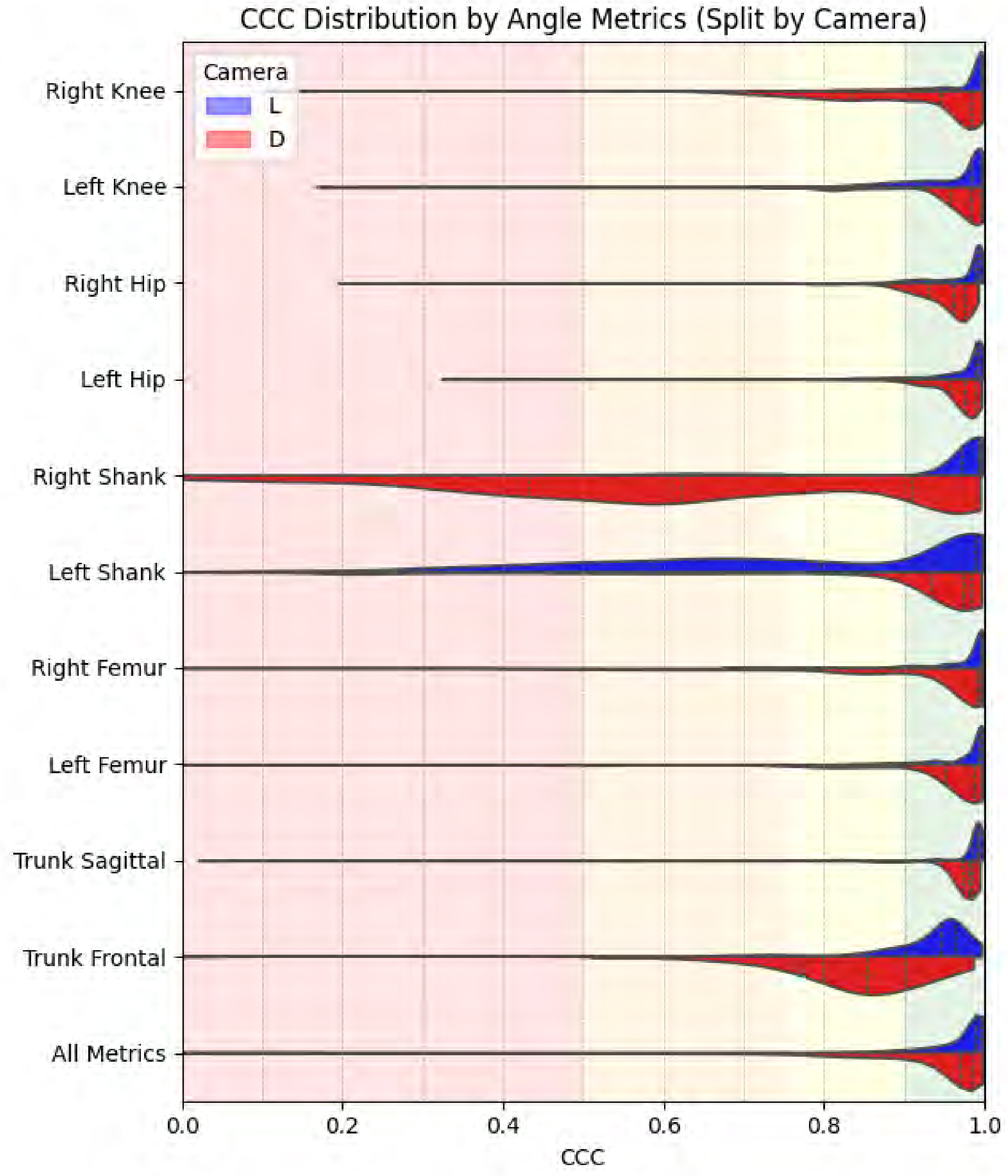


Figure 5.7 Density plot of the CCC values for every condition split by metric and by camera. The values were obtained using the 30 software iterations, where each pair of iterations gave a CCC value. It shows a generally tighter density function and a higher average value for the L515 than the D455. The background colour represents the evaluation of the reliability: <0.5 = poor, 0.5-0.74 = moderate, 0.75-0.89 = good, >0.9 = excellent.

Similarly to Figure 5.7, Figure 5.8 shows a violin plot of the density function of CCCs divided per condition instead of metric. We can see similar good performances for both cameras for conditions 01 to 04 with medians above 0.95, followed by weaker performances for conditions 10 and 11 with medians above 0.90. Conversely, both cameras do not perform at the same level for conditions 06, 08 and 09 where the L515 performs extremely well with medians above 0.98 but not the D455.

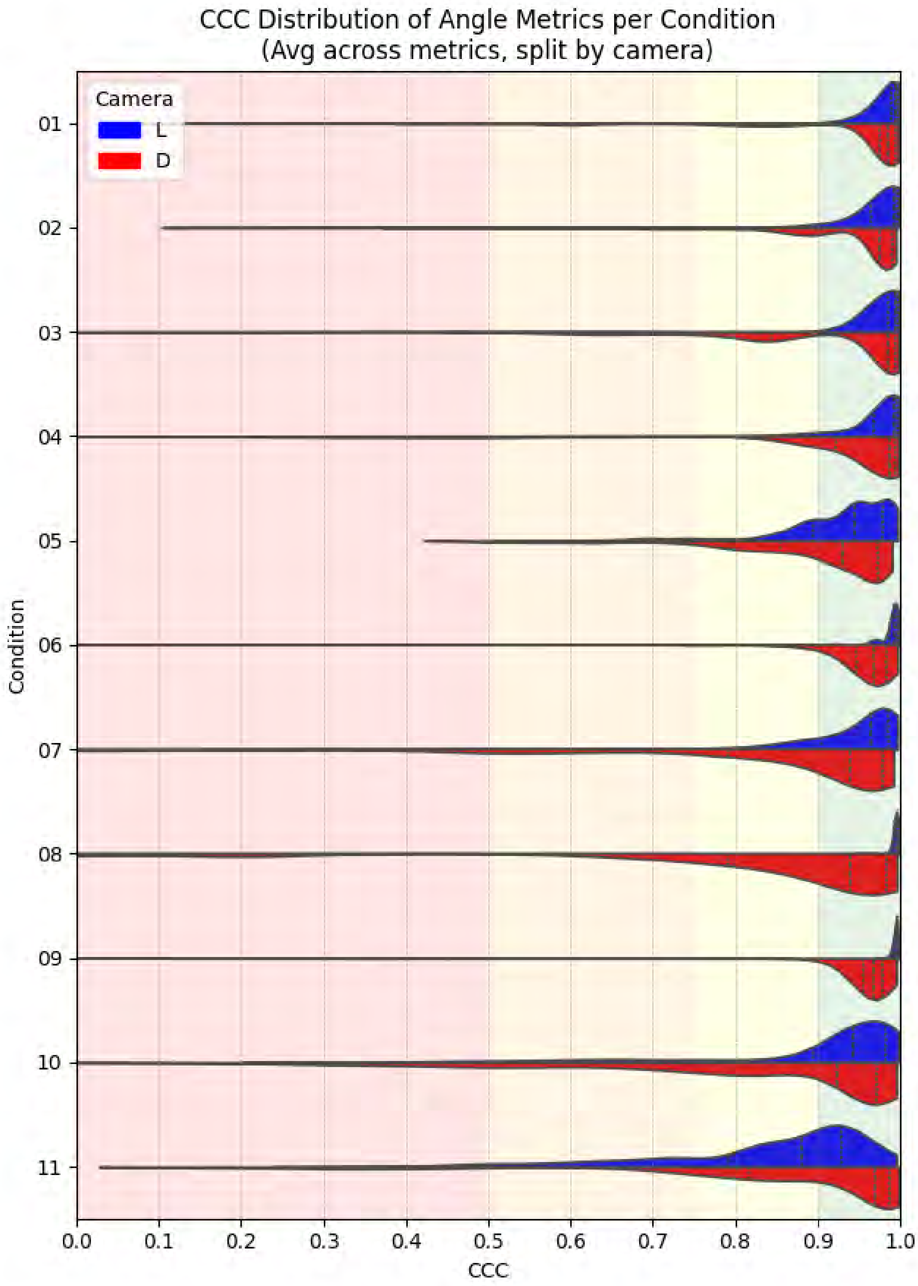


Figure 5.8 Density plot of the CCC values averaged for all metrics split by condition and by camera, similarly to Figure 5.7. Performance of the L515 and D455 vary between conditions, with the first four conditions showing similar performances, and the following conditions showing higher differences between cameras, generally favouring the L515 except for condition 11. The background colour represents the evaluation of the reliability: <0.5 = poor, 0.5-0.74 = moderate, 0.75-0.89 = good, >0.9 = excellent.

5.2.2 Precision of Velocity and Acceleration Metrics

The precision of the velocity metrics is very similar to the angle metrics. Figure 5.10 presents the CCC of the velocity metrics in the same way as Figure 5.7 and Figure 5.8. The CCC results are consistently lower and also have a larger spread for all metrics. Indeed, an excellent angle CCC will lead to an excellent velocity as well and a poor angle CCC will lead to an even lower velocity CCC.

This trend is shown in Figure 5.9 where all velocity lines have a peak left of the angle peak. The density value at the peak is also lower, showing a larger spread in the velocity data than the angle data. In general, the density peak for the velocity line is roughly half of the line for the angle data. The acceleration data has very low density and low average CCC, indicating very noisy data. All graphs show the same trend in acceleration data.

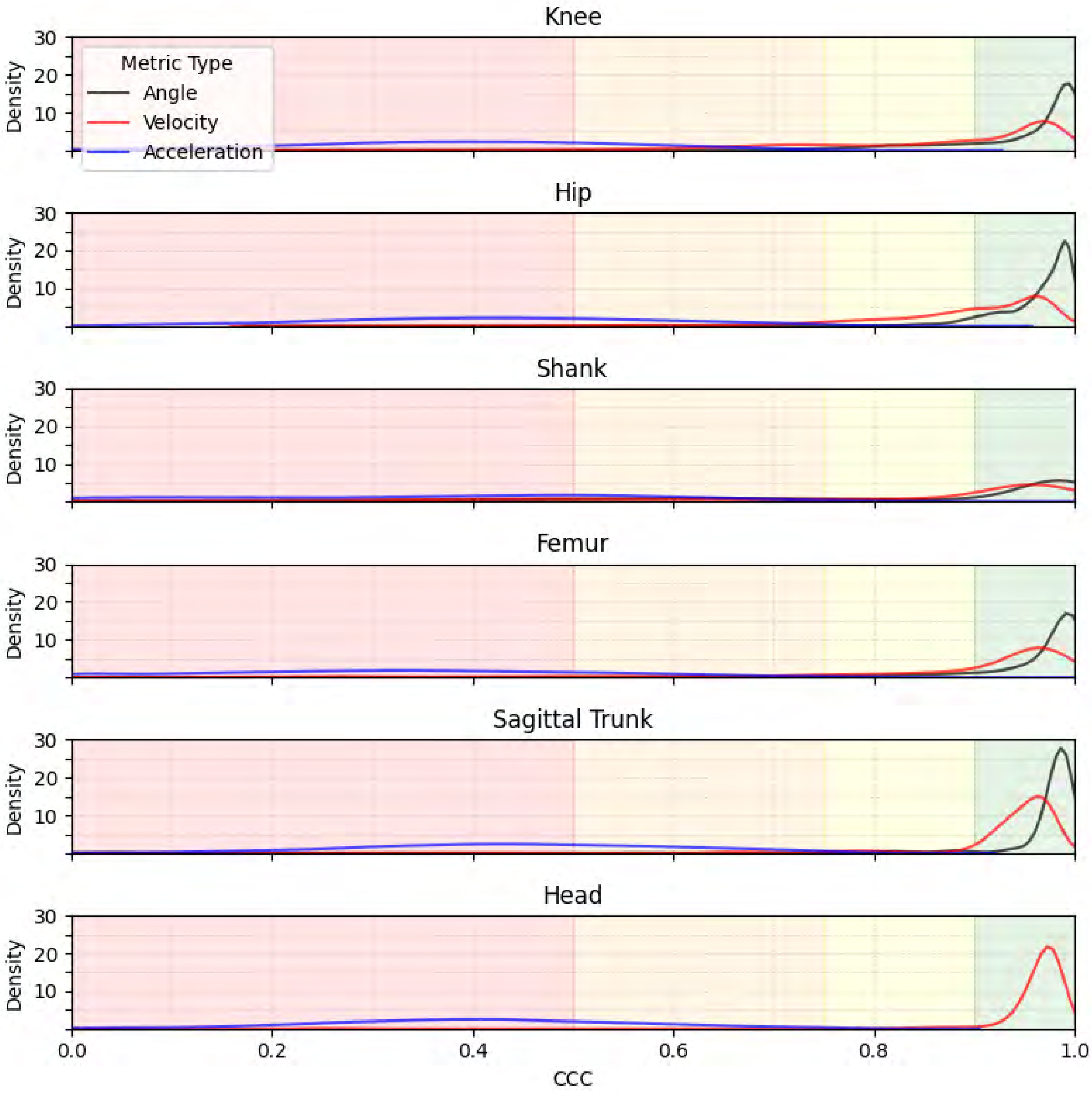
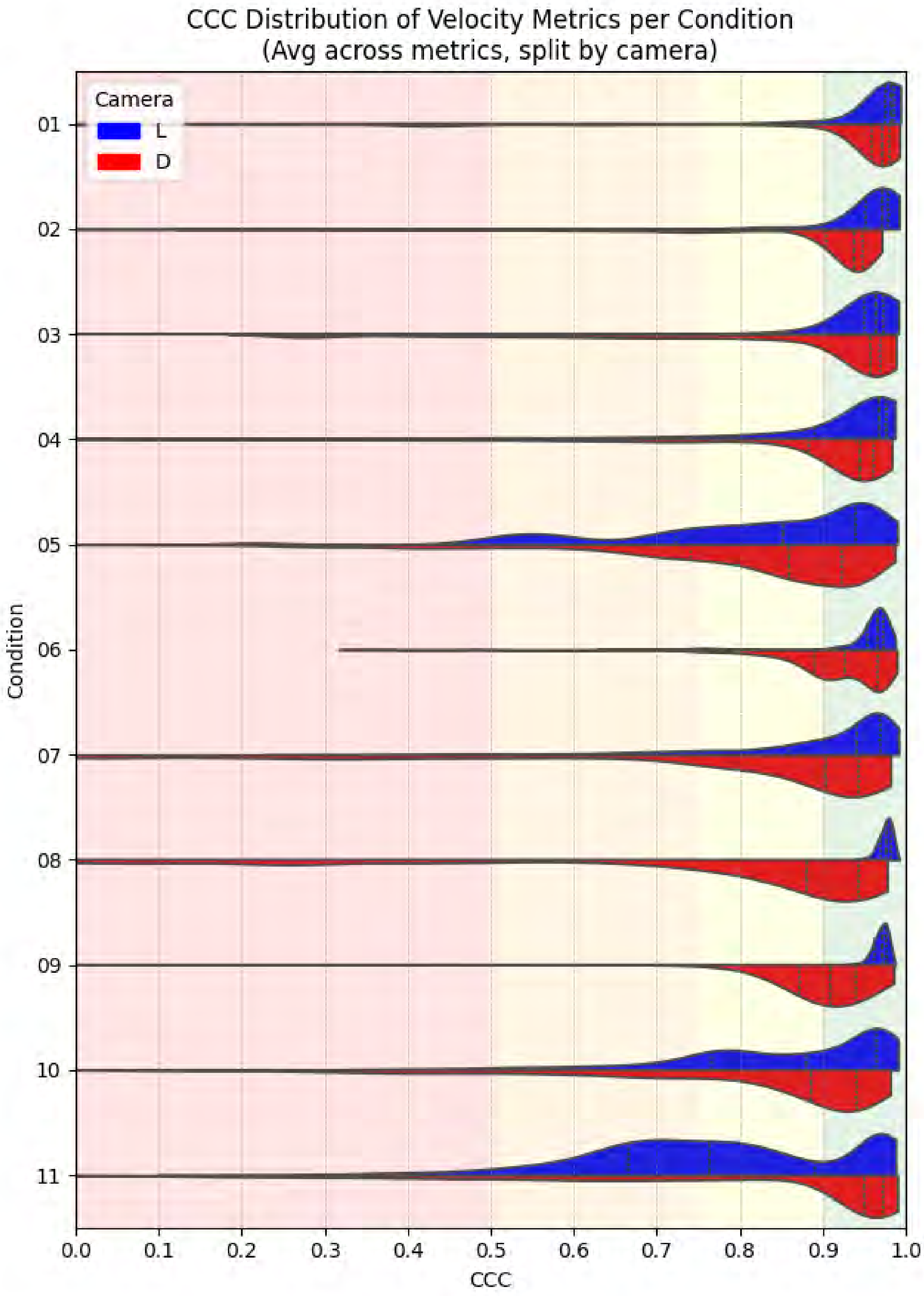
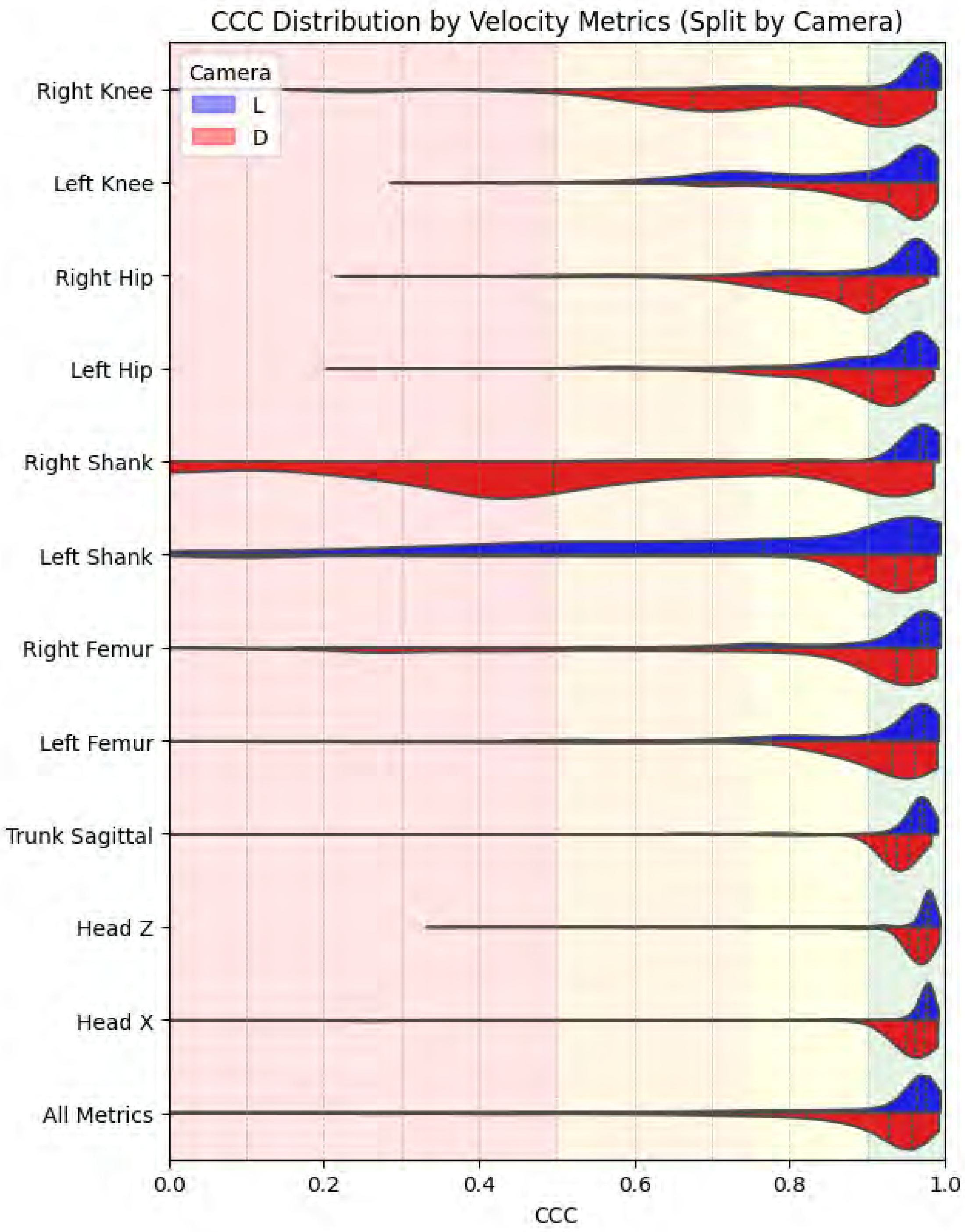


Figure 5.9 Density function of the CCC averaged for all conditions and both cameras. The figures present the six body parts from which metrics are computed. For example, the knee graph averages the data for both knees, both cameras and all 11 conditions; the blue line contains the angle data, velocity is in green, and acceleration in orange. There is no head angle, hence no blue line is present in the last graph. The background colour represents the evaluation of the reliability: <0.5 = poor, 0.5-0.74 = moderate, 0.75-0.89 = good, >0.9 = excellent.


CCC Distribution of Velocity Metrics per Condition
(Avg across metrics, split by camera)
Camera
L
D
Condition
01
02
03
04
05
06
07
08
09
10
11
CCC
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0

a)



b)

Figure 5.10 Density plots of the velocity metrics split by camera. The figure a) gives CCC values by condition while figure b) gives CCC values by metric. The general trends obtained from the angle metrics are also present here. The background colour represents the evaluation of the reliability: <0.5 = poor, 0.5-0.74 = moderate, 0.75-0.89 = good, >0.9 = excellent.

The last precision result related to velocity is shown in Figure 5.11 where the metrics of condition 08 filmed with the D455. The 95% CI shows that the interval is higher for the velocity than for the angle. The right shank angle shows a loss of position creating a very large peak in velocity for a short period of time, demonstrating this very common cause of error.

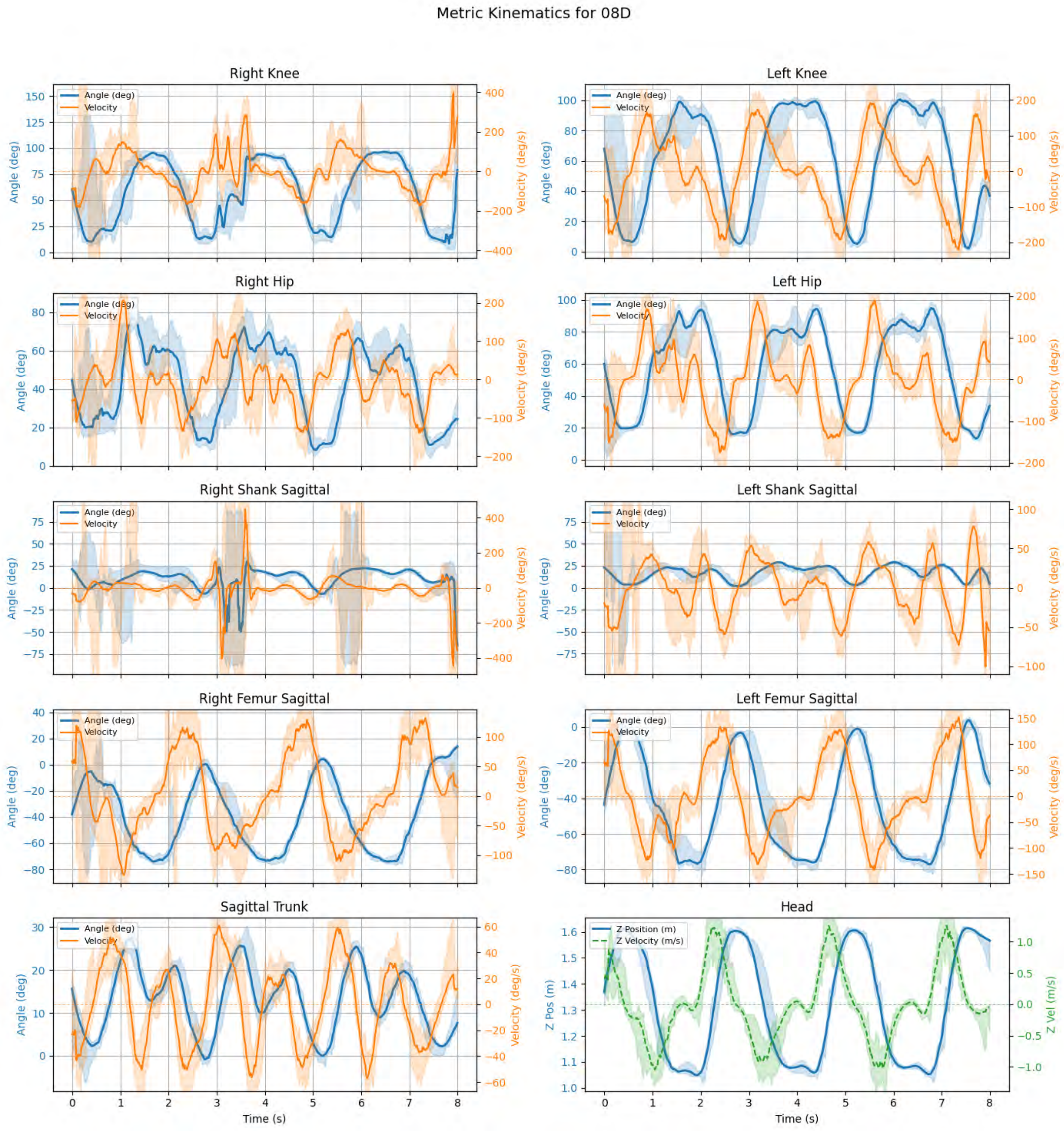


Figure 5.11 Comparison of the 10 metrics alongside their corresponding velocity curves. The shaded region represents the 95% CI for each curve, which is generally larger for the velocity than the original data. Errors in the position data cause very large spikes in the velocity data.

5.3 Accuracy Results

Accuracy results are obtained from comparing the experimental data to the ground truth. Metric variability is plotted for every condition in Appendix B, but general trends can be observed using condition 04 as an example in Figure 5.12. This figure shows that there is generally an offset between the ground truth and the estimated data, while the trajectory is generally well preserved. This figure shows the spread for the estimated data by plotting the 95% CI, but the Vicon ground truth falls outside of the CI most of the time. The last seconds of the recording tend to have a high occurrence of errors as can be seen with the left side metrics for the L515 data and for the right shank and knee angle for the D455. This type of figure gives a visual sense for the data that will be presented and analyzed with tables.

Table 5.2 presented the statistical results from the accuracy analysis to justify computing averages without the frontal trunk angles. Starting with correlation coefficients, the CCC values are much lower than the Pearson r values with the best metrics for both being the knee, hip and femur angles. The shank angles have a poor CCC and a moderate Pearson r. Finally, the sagittal trunk angle has the highest difference between CCC and Pearson r, going from 0.18 to 0.79. The MJAE, RMSE and PCEA are generally unrelated to the correlation coefficients, with the best MJAE and RMSE being for the shank angles, where the worst coefficients were obtained. The average MJAE is 23.4° and average RMSE is 26.5°, a value that is extremely poor, just like the 5.8% of PCEA@5°.

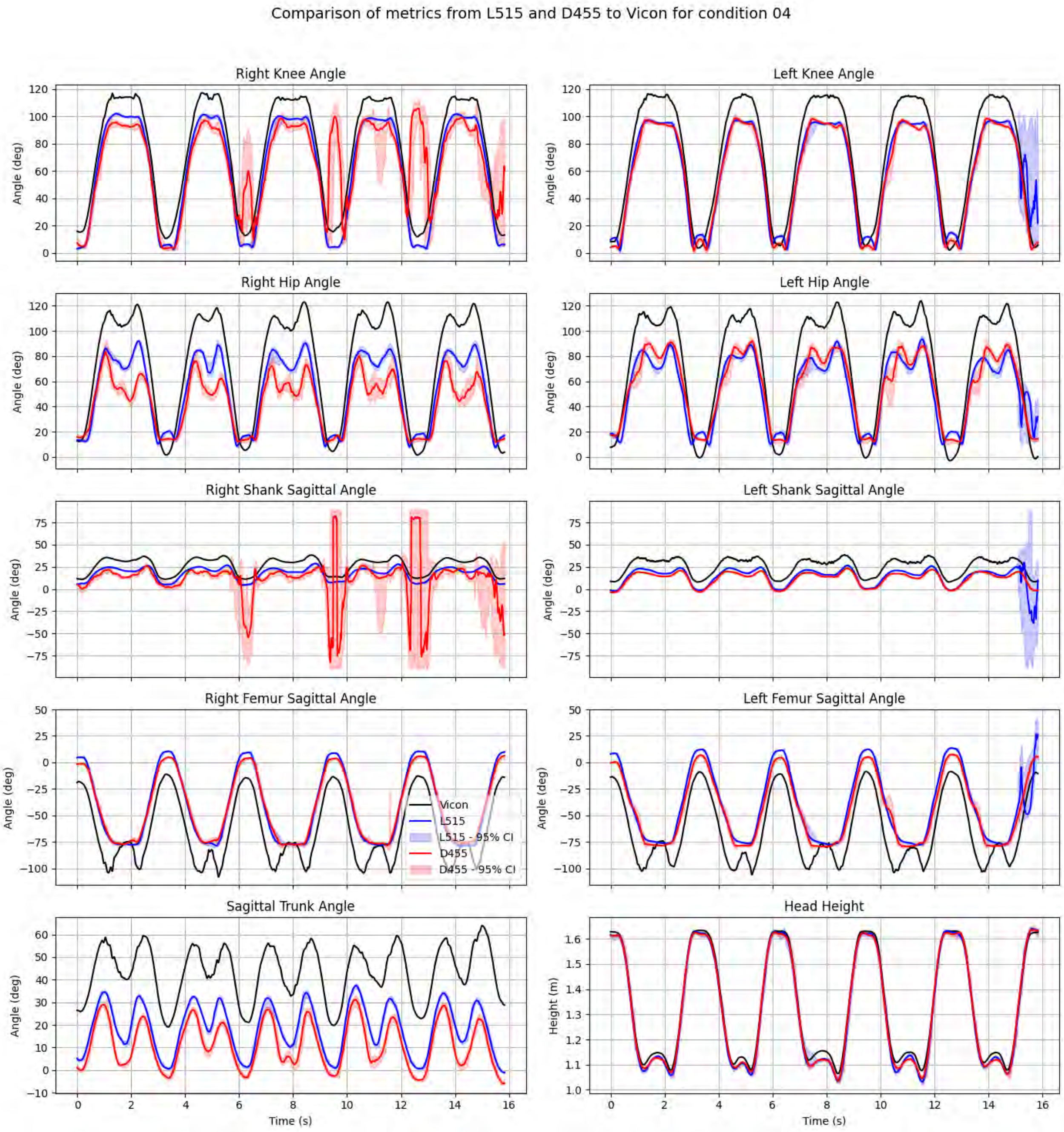
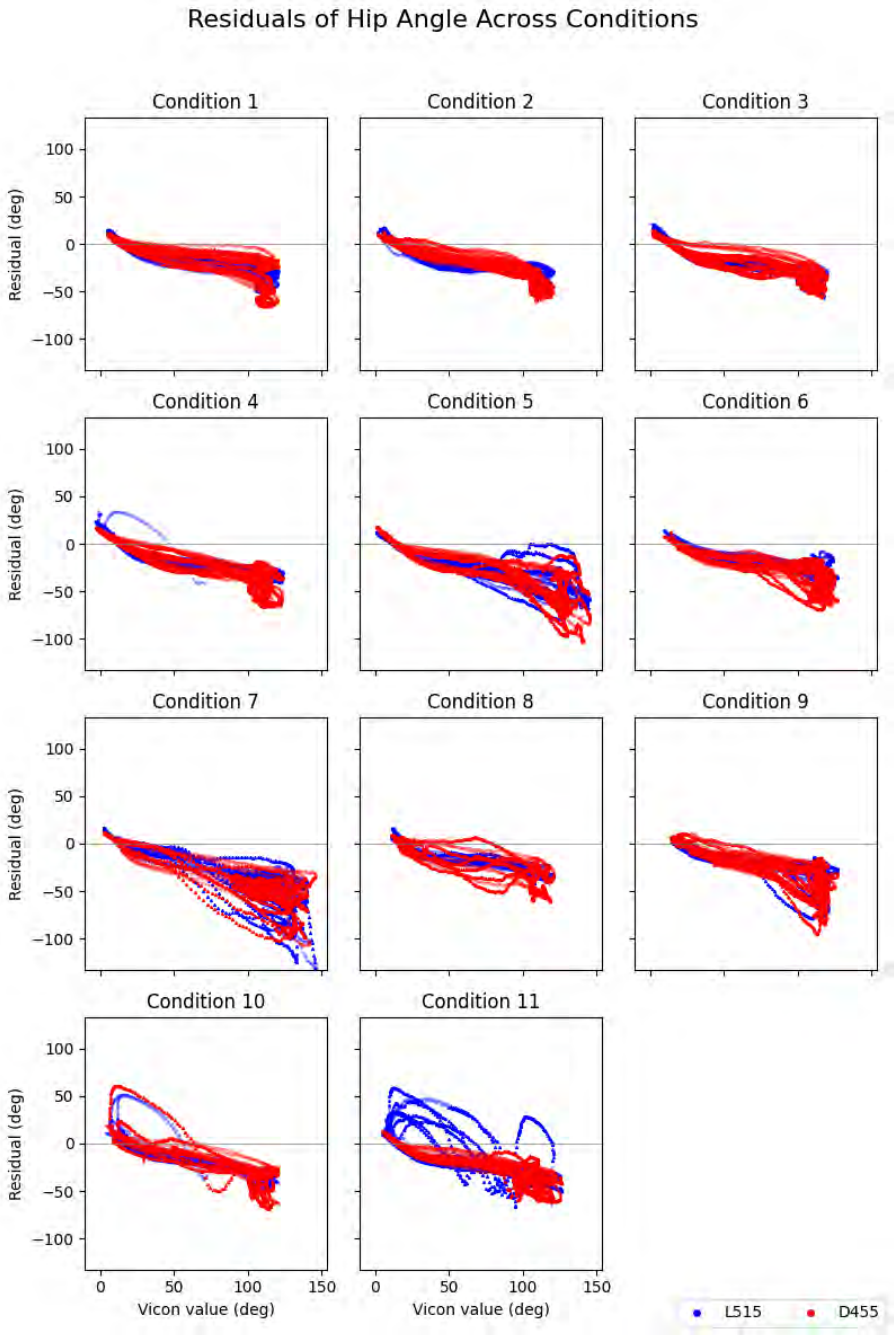


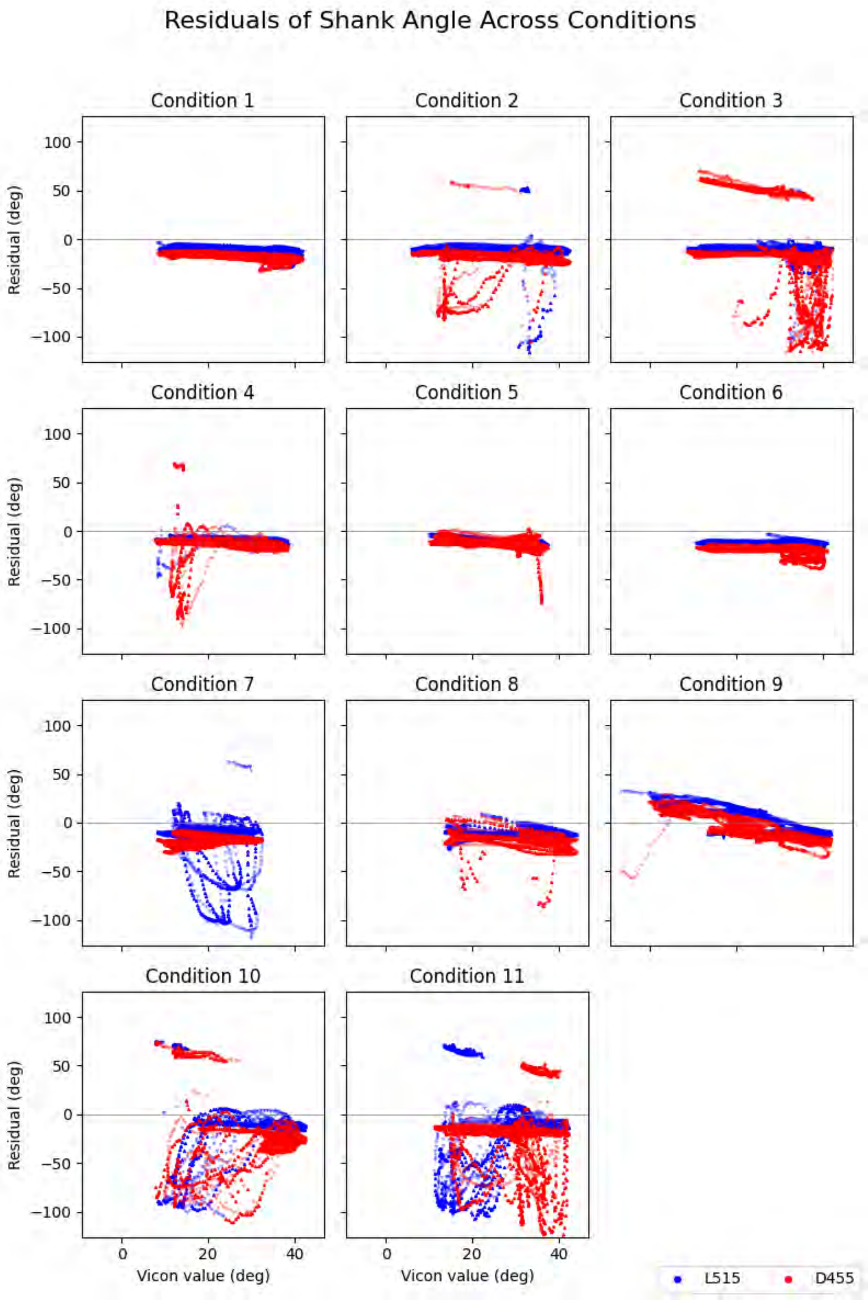
Figure 5.12 Comparison of all metrics with the Vicon ground truth for condition 04. The head height is plotted at the lower left to ease the interpretation of the curves with the StS movement.

5.3.1 Trajectory Trends of the Residuals

Apart from the systematic offset found in the trajectories of Figure 5.12, we can observe a systematic cyclic bias at different points of the StS motion for certain metrics. These trends are presented in Figure 5.13, where two examples of bias are shown in residual plots. These are the two extreme examples, where the hip angle error is inversely correlated to the magnitude of the ground truth and where the shank angle residual is flat for most conditions apart from the ankle position errors discussed in section 5.1.1. The residual plots for all metrics can be found in Appendix B.


Residuals of Hip Angle Across Conditions
Condition 1
Condition 2
Condition 3
Condition 4
Condition 5
Condition 6
Condition 7
Condition 8
Condition 9
Condition 10
Condition 11
Residual (deg)
100
50
0
−50
−100
0
50
100
150
Vicon value (deg)
L515
D455

a)



b)

Figure 5.13 Residual plots per metric for all conditions. The data is divided visually by colour according to the camera. Figure a) contains the residual of the hip angle and shows a non-constant relationship between Vicon ground truth and estimated data. Figure b) shows the residuals for the shank angle which are less dependent on the magnitude of the ground truth.

5.3.2 Symmetry Results

Asymmetry in StS can be caused by underlying issues of weakness, pain of range of motion, hence it is a characteristic that vision systems should be able to identify. Figure 5.14 shows the CCC between the right and left joint angle for the knee, hip, shank and femur. There are three points per condition: one for each vision system. The 95% CI is shown with error bars and represents the data obtained for the 30 software iterations of HPE for each recording. The L515 is generally more accurate than the D455, showing higher CCC values since most conditions were symmetrical. The shank angle for the L515 is problematic, as discussed in section 5.1.1. The only asymmetrical condition is 09 and the ground truth symmetry coefficient is 0.83 for the knee, 0.26 for the shank, 0.98 for the hip and 0.96 for the femur. The standard deviation between the ground truth and the L515 is 0.10 and 0.20 for the D455 for condition 09.

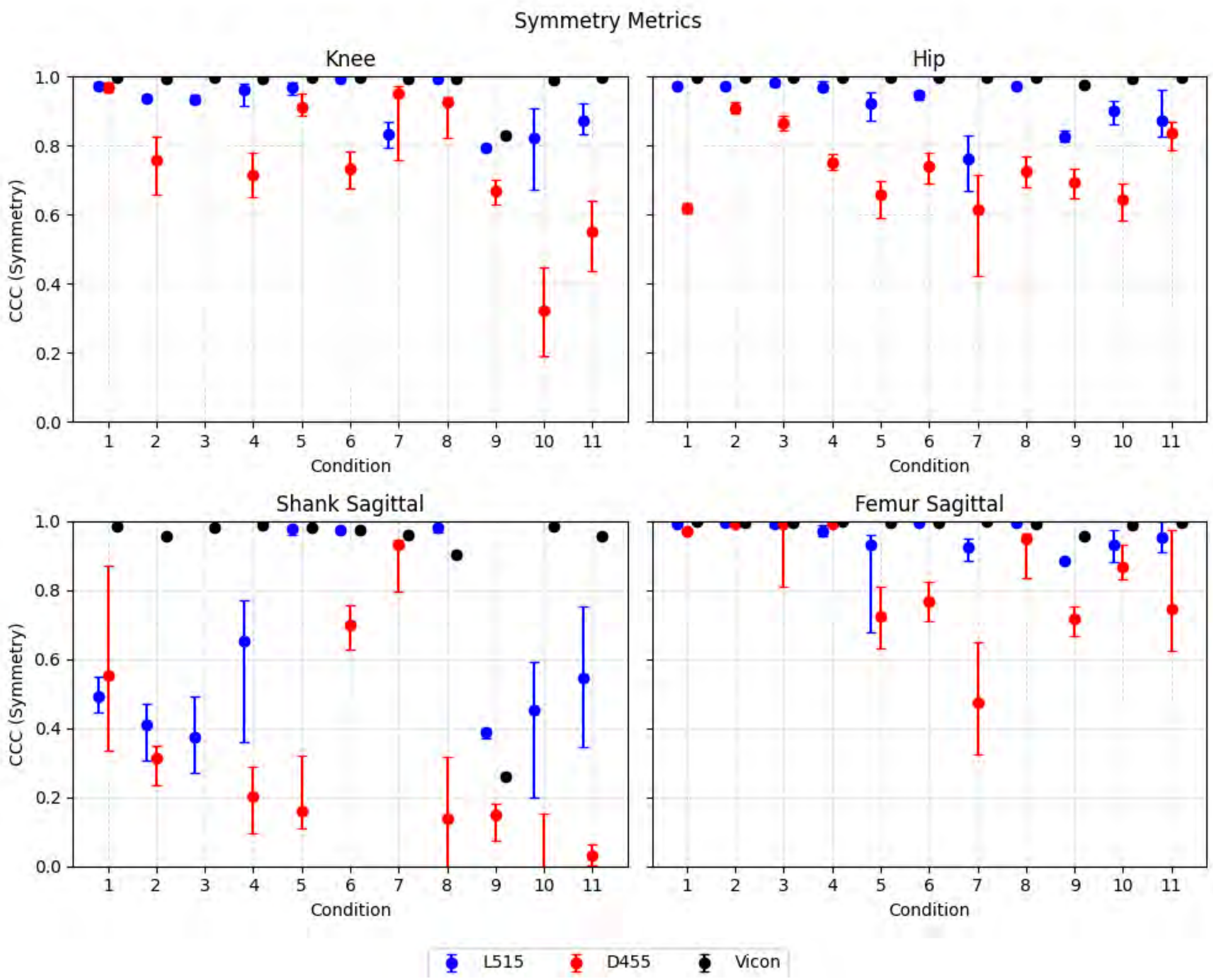


Figure 5.14 Symmetry plot for the knee, hip, shank and femur where the right and left angles are compared with CCC.

5.3.3 Velocity Results

The accuracy results for the velocity metrics are shown in Figure 5.15. Peak velocities for each StS repetition are identified and averaged, leaving the maximal angular velocity, positive and negative for each metric. Note that some data points for the shank velocities are not shown in the figure, since they have residuals higher than 100%.

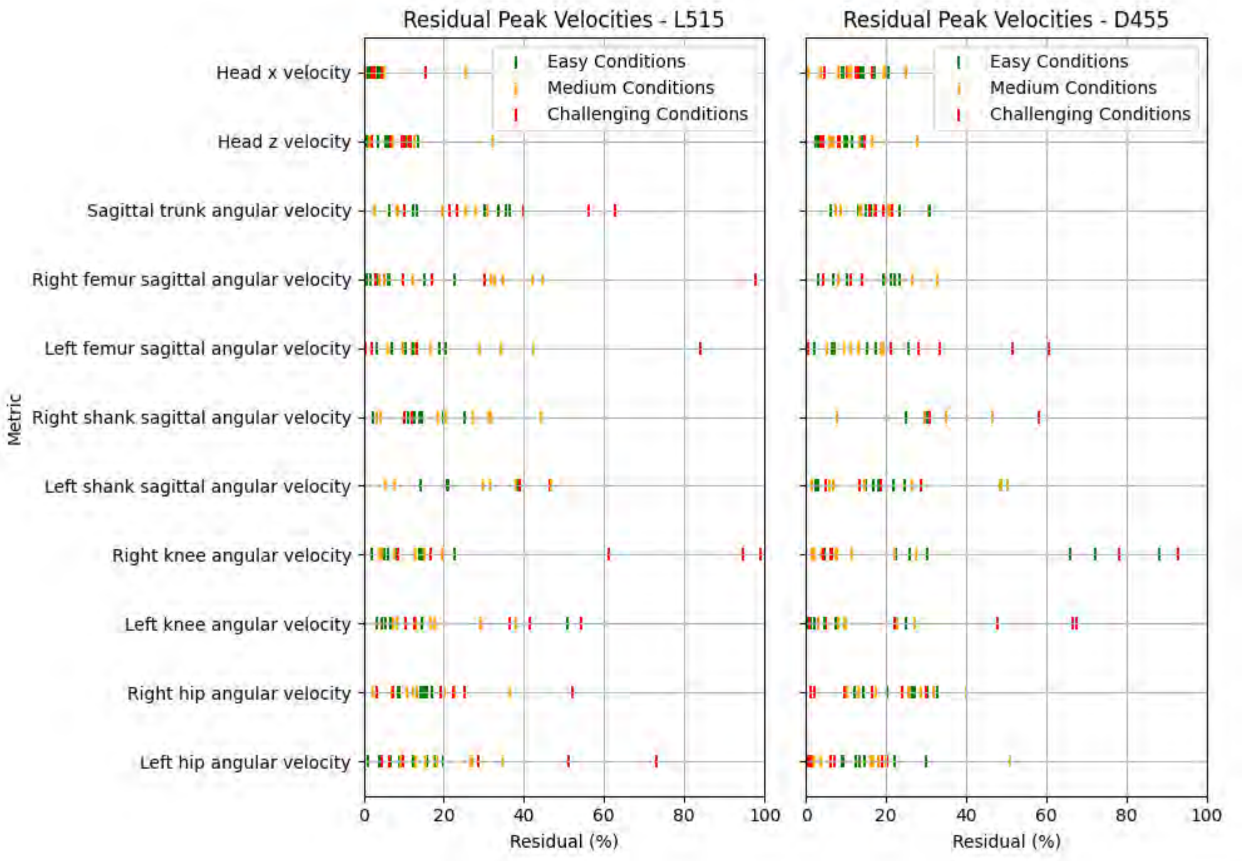


Figure 5.15 Residual of the positive and negative peak velocities averaged for all conditions and separated by metric. The three groups of condition difficulty are differentiated by their colour. There is no obvious difference between both cameras, with some metrics performing better for the L515 and some for the D455. The largest residuals are found for the challenging conditions. All residuals for the easy conditions are below 40% except for the knee velocities.

5.4 Results with Offsets

After obtaining the results, a systematic bias was found for most joints. An analysis of the accuracy with a systematic offset, per joint, is described in the following section.

5.4.1 Methodology of Computing the Offset

Offsets are computed across all conditions, for every joint type, per camera. This means that the offset for the left and right knee will be the same for the L515 but there will be a different offset for the left and right knee measured with the D455. The results vary slightly for left vs right metrics, but applying a general offset seems more reasonable for truly evaluating the HPE system based on the assumption that the angles are measured the same way regardless of the laterality.

The average of the metric angle residual was calculated to find the value of the offset, but first some of the data was removed. The highest 25% of residuals and lowest 25% were both ignored, leaving the middle 50% of the residual data to compute the average. Equation 6.1 is the offset equation, with $\mathcal{R}_{25-75}$ being the middle 50% of residuals and r being the residual for a single frame.

$$\Delta\theta = \frac{1}{|\mathcal{R}_{25-75}|} \sum_{r \in \mathcal{R}_{25-75}} r \qquad 4.1$$

This method is more robust to the multiple outliers found in some of the metrics, which would have polluted the average. The 50th percentile was used exclusively for this calculation, since the 2.5th and 97.5th percentile contain more outliers. The offsets calculated are presented in Table 5.4. The CCC between the L515 and D455 offset is 0.97, showing excellent agreement, but the difference is still larger than 3° for most metrics.

Table 5.4 Offset calculated per joint and per metric. The largest offset calculated is for the sagittal femur angle followed by the knee angle.

Metric	L515 Offset	D455 Offset
Femur Angle	24.4	20.8
Hip Angle	-25.1	-25.0
Knee Angle	-16.4	-19.6
Shank Angle	-10.4	-16.8
Sagittal Trunk Angle	-25.5	-31.7

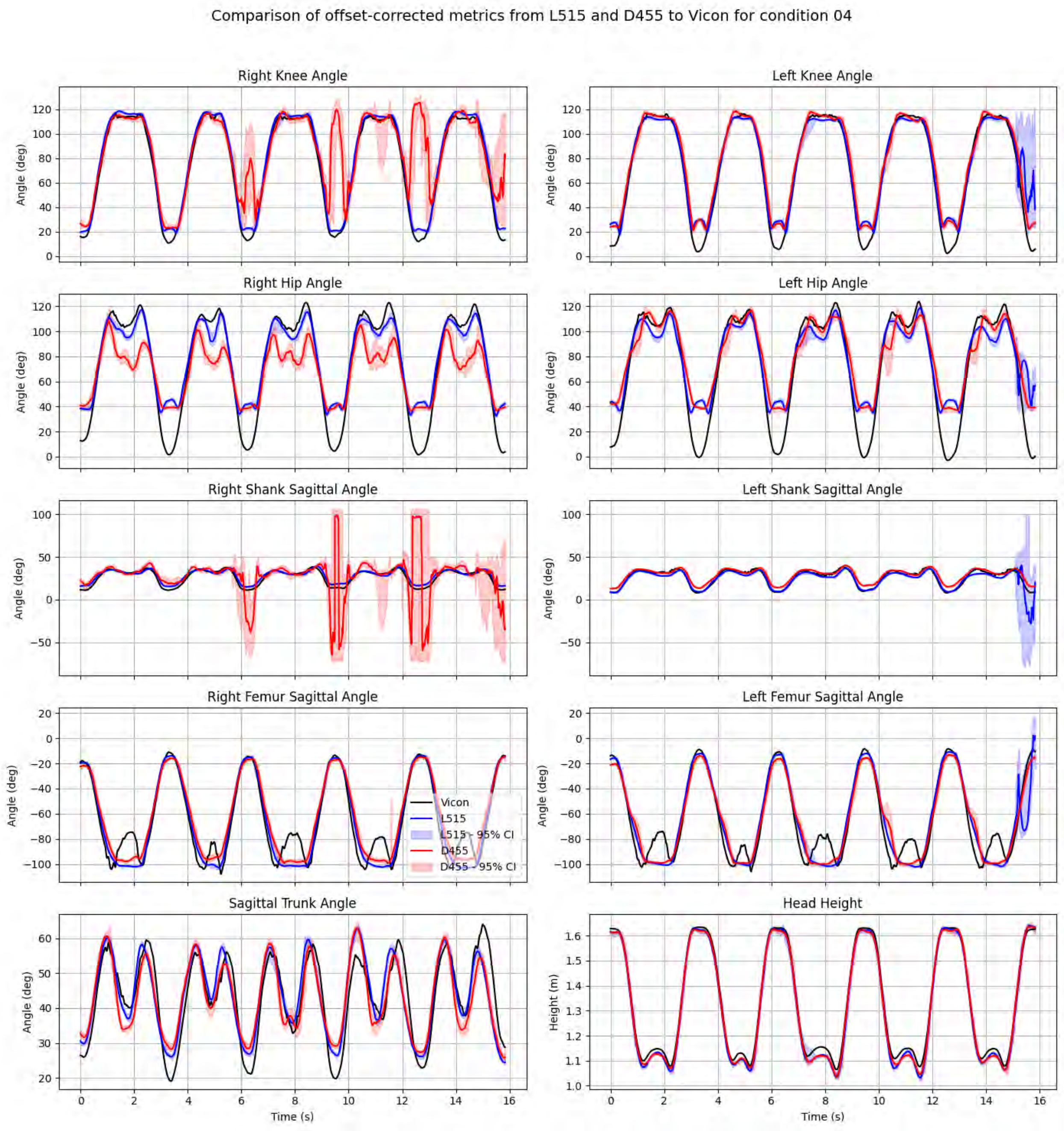
5.4.2 Results from the Corrected Data

5.4.2.1 General Statistics and Figures

The same data evaluation methodology is applied to the corrected data as was applied to the initial data. Table 5.5 gives the statistics results from the corrected data per metric. When compared to the results from the initial, non-corrected data in Table 5.3, a few differences can be observed. All statistics improve, except for the Pearson r (which is blind to absolute differences), with average CCC going from 0.53 to 0.75, MJAE going from 23.4° to 11.2°, RMSE going from 26.5° to 15.6° and percentage of correctly estimated angles within an angle error, PCEA@3° to PCEA@15° spreading from 3.3%-31.9% to 30.8%-78.0%. The best CCC after correction are for the knee, hip and femur almost equally. The sagittal trunk angle has the highest CCC improvement going from 0.18 to 0.72 and the right shank shows the worst correlation performance.

Table 5.5 Statistics of the corrected data per metric.

Metric	CCC	Pearson r	MJAE (°)	RMSE (°)	PCEA@3	PCEA@5	PCEA@15
Right Knee Angle	0.78	0.8	10.9	16.6	34.1%	51.0%	84.2%
Left Knee Angle	0.86	0.91	10.5	15.2	29.3%	45.6%	79.8%
Right Hip Angle	0.79	0.92	16.1	20.1	12.3%	20.9%	56.7%
Left Hip Angle	0.82	0.92	15.3	19.6	13.9%	23.2%	60.9%
Right Shank Sagittal Angle	0.55	0.59	11.1	16.6	49.4%	65.9%	86.0%
Left Shank Sagittal Angle	0.58	0.66	7.9	12.4	49.8%	67.9%	85.9%
Right Femur Sagittal Angle	0.83	0.85	11.1	15.7	29.1%	44.1%	79.1%
Left Femur Sagittal Angle	0.81	0.84	11.3	15.9	25.8%	40.9%	78.2%
Sagittal Trunk Angle	0.72	0.79	6.7	8.4	33.9%	53.6%	91.5%
All Metrics	0.75	0.81	11.2	15.6	30.8%	45.9%	78.0%


Comparison of offset-corrected metrics from L515 and D455 to Vicon for condition 04

Right Knee Angle

Left Knee Angle

Right Hip Angle

Left Hip Angle

Right Shank Sagittal Angle

Left Shank Sagittal Angle

Right Femur Sagittal Angle

Left Femur Sagittal Angle

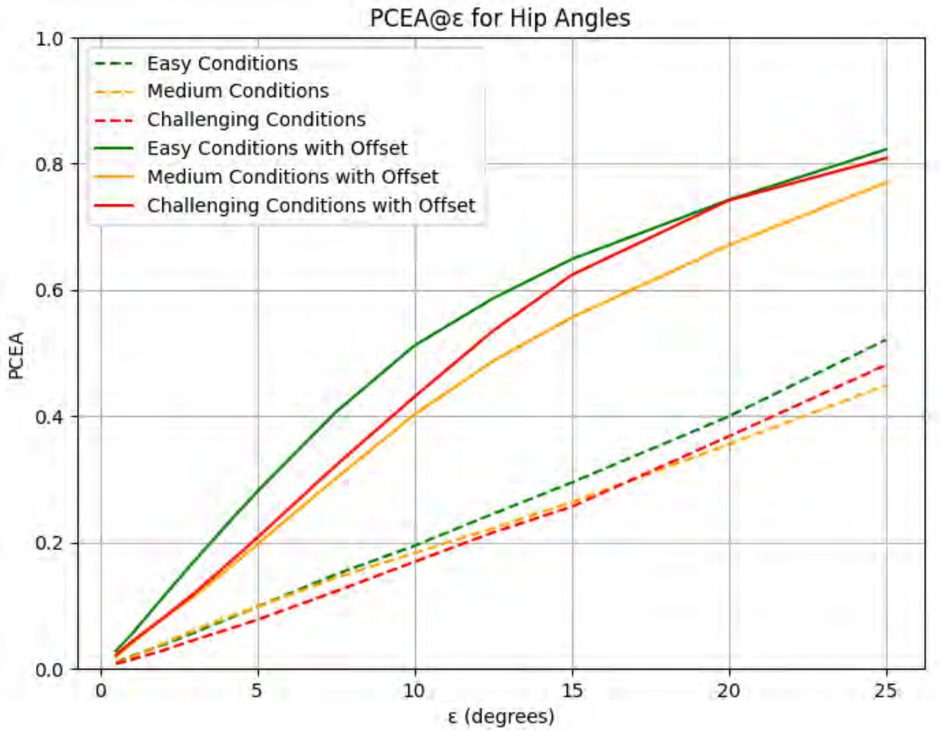
Sagittal Trunk Angle

Head Height

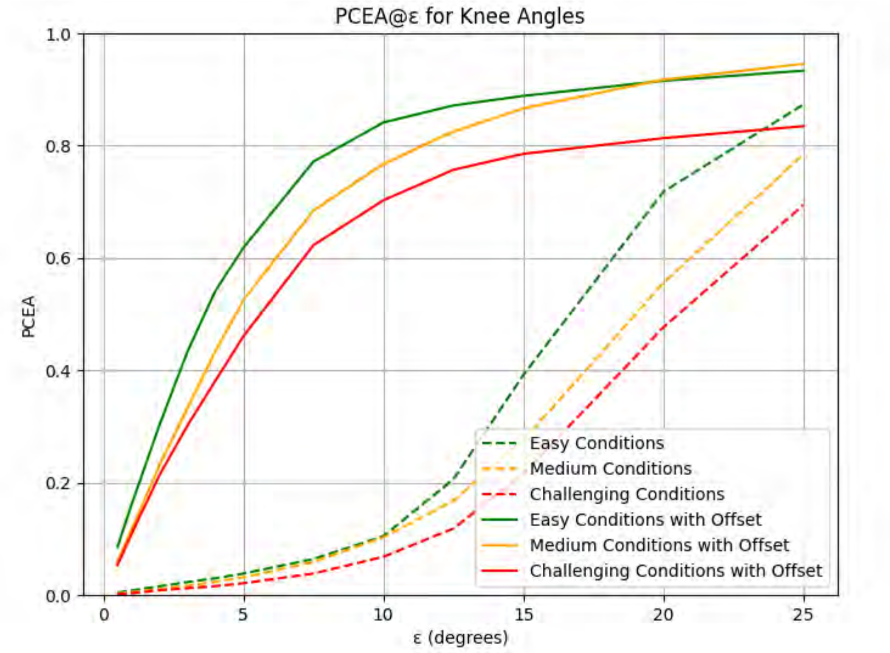
Figure 5.16 Comparison of the corrected metrics with the Vicon ground truth for condition 04 using the same layout as Figure 5.12. The trajectories overlap generally with extreme data at full extension of full flexion still causing problems.

5.4.2.2 Percentage of Correctly Estimated Points Results

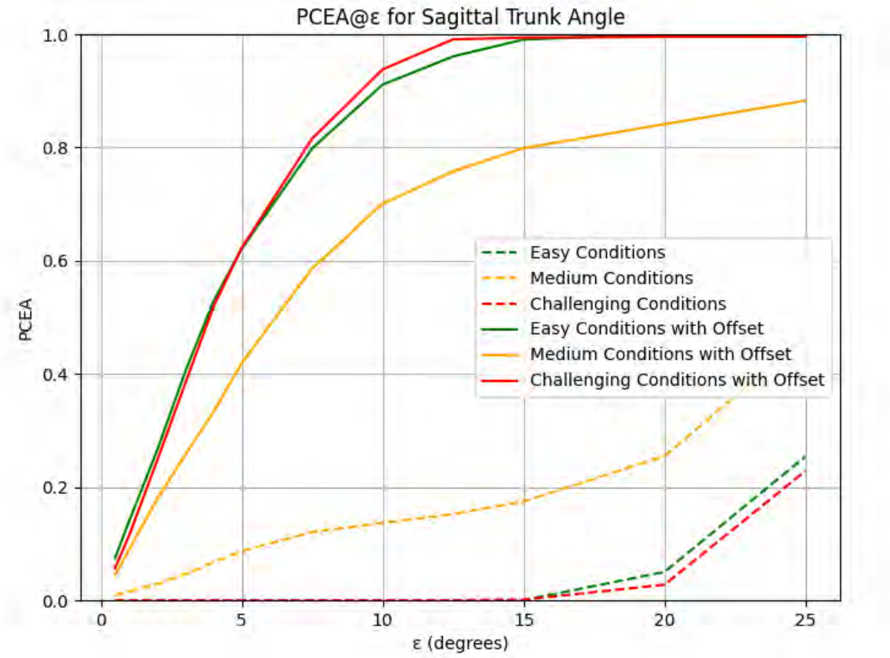
PCEA is very sensitive to offsets since it only relies on absolute values. Figure 5.17 shows the difference between corrected and non-corrected data, for all metrics. It shows that the PCEA for all metrics largely improves after correction, but not equally for all metrics. Figure 5.17 shows the percentage of correctly estimated angles within a certain angle error ϵ (PCEA@ϵ) for three examples of joints. The results are divided by metrics and combined for both cameras and by condition difficulty. The corrected PCEA curve for the hip angles is almost linear, while the knee and sagittal trunk increase quickly, followed by a plateau.



a)



b)



c)

Figure 5.17 PCEA@ϵ for three example metrics that are affected differently by the offset. The colour code shows a difference in difficulty of the condition with green being the easiest and generally having the highest PCEA and red being the most challenging and generally having the lowest. Figure a) shows the hip angle, which is helped marginally by the offset, b) the knee angle, which is helped by a moderate amount by the offset and c) the sagittal trunk angle which is helped a lot by the offset.

CHAPTER 6 DISCUSSION

6.1 Interpretation of Qualitative Observations

The qualitative issues described above highlight specific limitations of the low-cost markerless motion capture system that must be considered when interpreting the quantitative results. First, the frequent loss of foot tracking, especially on the right side with the D455 and the left with the L515, indicates a vulnerability of the pose estimation algorithm in maintaining full-body consistency in upright positions. The timing of the loss of position indicates that the algorithm might be sensitive to lighting or to the standing position. The loss of a single joint affects two metrics in this case: the shank sagittal angle and the knee angle, since the knee angle is also based on the ankle position. In the case of 5StS, the foot position could be assumed to be static, since there is constant contact with the ground for both feet. This fix neutralizes this issue for StS evaluation, but the system would remain exposed to this issue during other types of movements or in StS scenarios with patients unable to keep their feet still.

The loss of this joint centre is not symmetrical for both cameras, with the L515 losing the left side more often and the D455 losing the right more often, as seen in Table 5.1. This discrepancy is surprising given that both cameras viewed the scene from almost the same angle, minimizing the differences in occlusion and lighting. The local characteristics of the scene in conjunction with the particularities of each camera gave two very different results across the conditions. However, since the errors consistently appeared on one side for each camera, with no intermediate or mixed behaviour, additional experimental conditions would be needed to better understand what triggers these side-specific failures.

This problem is most pronounced for conditions 10 and 11 when loose-fitting shorts obscured most of the thighs and partially the knees but had no impact on the visibility of the feet. It suggests that the issue is not only with the foot itself but is also related to the whole leg, where more obstruction on the upper leg seems to lower the accuracy of the estimates for the lower leg.

As explained in section 4.5.2, the joint centre data for the Vicon system was problematic for the hip joints due to obstructed markers. While the calculation of joint angles by the Vicon software was generally more robust to such errors, artifacts still appear in the trajectory of the hip and femur angles. Figure 5.4 illustrates two examples of this unusual trajectory. The main difference between these conditions is the foot placement and the extent of sagittal trunk movement, which hints that the angle computation error may stem from inaccuracies in trunk or hip estimation during movements with large ranges of motion. Thus, larger femur angle discrepancies observed in conditions involving significant trunk flexion may be partially attributable to errors in the reference data.

6.2 Interpretation of Precision Results

6.2.1 Angle Metrics

Table 5.3 shows interesting results that can be coupled with the metric variability figures in Appendix B. The first visible trend from the table is the differences between the conditions. This discrepancy shows that the system behaves more repeatedly for easy scenarios than for the more difficult scenarios. This does not mean that easier scenarios are more accurate, but simply that repeated interpretations of a recording are consistent.

Conditions 05 and 07 involve an exaggerated trunk flexion and show some of the worst CCC values for hip angles and femur sagittal angles. These metrics depend on accurately localizing the hip joint, which becomes more difficult when the trunk leans forward. A forward trunk posture reduces the number of pixels vertically aligned with the torso, diminishing the visible surface area available for pose estimation and increasing the reliance on noisy data. The hip joint itself is anatomically challenging to estimate from the front when seating, due to the surrounding soft tissue and the high mobility of the spine and the pelvis without obvious external displacements. An acute hip angle accentuates these limitations by occluding more of the lower abdomen, leading to worse hip joint estimations, as shown in conditions 05 and 07.

CCC density functions shown in Figure 5.7 allows for an analysis of correlation per metric and per camera. While the averages in Table 5.3 showed a modest advantage for the L515, the density functions tell a more nuanced story. For most metrics, the spread for all metrics except the left shank shows a larger spread for the D455, not only with min-max range but also with the first and third quartiles. This shows that the interpretation of the D455 data is less repeatable than the L515 data. This could mean that the ToF data is more information rich than the stereo data and does not require the pose estimation to estimate as much, introducing variability on top of the same recording. In general, the precision data is excellent with CCC above 0.90. For moderately challenging scenarios the L515 performs better, but for challenging scenarios there is no difference between cameras.

While precision results are valuable as a first step and give valuable insight into the performance of the system, the accuracy results give a much better evaluation of how trustworthy the product can be.

6.2.2 Velocity and Acceleration Metrics

Velocity and Acceleration are less precise than the angle data they are computed upon. The lower apparent noisiness of the velocity data is partly due to the 20-frame moving average window for velocity analysis. This window serves an important purpose in keeping the CCC values higher by removing noise that is not problematic in the angle data but that would become problematic after derivation. The acceleration is computed directly from the velocity data without a second averaging step and is extremely poor. The acceleration as computed is certainly not sufficiently reliable, hence results related to acceleration are presented in the appendix but will not be further discussed.

The head velocities have the highest CCC results for the velocity, with the Z metric (height) outperforming the X metric (depth) slightly. The X metric is more reliant on the depth data while the Z data is more reliant on the RGB data. This shows that the detection of the head in depth is less precise than the detection in the RGB image. In contrast, while the shank velocity in Figure 5.8 appears visually closer to the Vicon position data, a closer inspection of the y-axis scale reveals that the data density is similarly poor for both types of metrics.

6.3 Interpretation of Accuracy Result

6.3.1 General Statistics and Figures

The results from Table 5.2 show that while the precision in the trajectory is generally good at $r = 0.81$, the average CCC is only moderate at 0.53. This difference shows that the CCC could be improved by a better accuracy. A large offset can be found in Figure 5.12 for most metrics with the largest being for the sagittal trunk angle. This angle also has the highest difference between CCC and Pearson r, confirming that the lower CCC is caused by the offset to the ground truth. This offset is problematic but can be corrected through systematic adjustment. Section 6.4 explores this additional processing step in more detail.

A different observation in Figure 5.12 is the higher incidence of errors in the last second of the recording. The last repetition was immediately followed by a forward walk, where one foot was lifted shortly after full knee extension. This might have altered the posture and symmetry of the final movement, leading to inconsistencies that challenged the pose estimation algorithm. This trend in error is unlikely to result from proximity to the end of the recording, as the pose estimation algorithm was applied before the initial trimming that removed a few seconds from both the beginning and end of the sequence. It might indicate that the HPE has not been trained as well on off balance movements as on the standard symmetrical StS movements.

6.3.2 Analysis of Residual Trends

Residual figures give a good understanding of where in the movement the error occurs and how much is caused by noise versus a varying bias. The residual figures are presented per metric with one plot per condition to showcase the repeatability of the error. A constant offset appears as a flat line of points in this type of graph, and this is what we see for the sagittal shank angle. It means that there is no additional bias at a specific point of the movement for the shank.

The knee angle and the femur angle behave differently by having a constant bias through the range of motion, except at a low angle. For certain conditions, the complete extension of the knee is underestimated. This is a surprising finding, since with a frontal viewpoint, the last degrees before extension have a large impact on the forward or backward position of the knee joint. This effect is related to the small-angle approximation in engineering, since a fully extended knee joint in s standing position moves almost entirely horizontally Hence, when measuring its position, the angle can be measured with even more accuracy since small changes of the angle, results in a large displacement. This is the opposite of what the data shows, with the knee angle reaching full extension with Vicon but stopping before full extension with HPE. This might show a bias against full extensions for HPE, as we can also see in the hip angle.

The residual figures for the hip angle are also interesting by showing a large varying bias: overestimating small angles and underestimating large ones. Just like for the knee, the extension part of StS is not well captured at the hip, but the seated position is also consistently mischaracterized. The hip angle when sitting is noisier than when standing but is always underestimated compared with the rest of the movement.

Knowing the general trend of residuals allows to visually determine where in the movement the angles are more precise and where they cannot be trusted. Unlike with all other statistics that average the whole waveform, we can see that femur angles for the standing half of the motion, for example, are very precise.

6.3.3 Analysis of Symmetry

One notable result is that the CCC value for condition 9 is not the only low one for the L515, but for any CCC below 0.95 the 95% CI is the smallest by far, meaning that the asymmetry is measured repeatedly instead of being caused by outliers. The inability of the D455 to match the left and right metrics for symmetrical movements is problematic. It means that the data it gathers is dependent on laterality or that the signal-to-noise ratio is too high for the measurement to be repeatable.

6.3.4 Analysis of Velocity

The velocity results are very susceptible to any outlier in their input data, making the results highly varied depending on the presence of outliers or not. The velocity of the head is the most accurate on average similarly to the precision results.

When looking at the density of points in Figure 5.15, we can see the peak between zero and 20, with a rapid decrease followed by a constant density between 50% and 100%. This plateau can be due to the susceptibility to outlier data of the peak finding function. The velocity peaks are identified digitally according to their timing and their prominence, hence when an outlier causes a large spike in the data, a detection is triggered. The detection algorithm then enters a sleep stage, in order to exclusively acquire one peak per repetition, but the algorithm starts looking for peaks 15% before the full StS period, to account for slight variation in timing. This 15% tolerance in time can accumulate throughout the recording and the outlier points can be the only peaks recorded.

6.4 Further Analysis of Offsets

The offset found in the evaluation of the metrics is surprisingly consistent per metric and camera. This quasi-constant difference could indicate that the angles are not calculated in the same way by Vicon and Nuitrack or that there is a systematic bias introduced in the ground truth data or in the estimated data. Bias linked to the ground truth has been discussed at the beginning of this chapter and will not be explored further. A bias linked to the pose estimation algorithm is highly likely, since the algorithm generalizes the skeletal structure of the subject based on soft tissue. There is high variability in body types and musculoskeletal appearance that can be lost in an RGB-D video, or simply not present in the training data.

We can see sections of the trajectory for certain metrics that have a systematic error, such as the full extension of the knee angle, but the easiest error to fix is a constant offset. While it is debateable whether the constant offset could also be applied to a different subject, adjusting for this offset allows us to evaluate the resulting corrected metrics with higher details. This analysis step was not predicted during the development of this study; hence the methodology, results and discussion of this step will be presented in this section 6.4.

6.4.1 Interpretation of the Corrected Results

When compared to the results from the initial, non-corrected data in Table 5.3, a few differences can be observed. First, the large discrepancy between CCC and Pearson r is reduced, as expected. The largest difference between CCC and Pearson R is at the hips for the corrected data where it looks like the trajectory of the hips has a scale issue, not completely reaching the maximum nor the minimum of the waveform.

Second, we can see that the correlation coefficients are generally inversely proportional to the angle errors statistics. Indeed, the hips have the best Pearson r of 0.92 but the worst PCEA, MJAE and RMSE at all levels. Conversely, the shank angle has the worst CCC and Pearson r, but the best PCEA@3 and @5, as well as among the best MJAE, RMSE and PCEA@15. The scale of the data seems to play a role in this observation; indeed, the correlation coefficients observe proportional changes in the measurements, where the other five metrics evaluate absolute differences. As a result, angles that have a large range of motion such as the hip, then show excellent correlation, but low absolute agreement. We can see this behaviour elsewhere as well, as the highest difference between CCC and Pearson r ($CCC - r = 0.115$) also happens at the hips. It shows that despite the correction offset, the accuracy component of CCC is harder to fulfill with low absolute agreement.

When comparing MJAE and RMSE, the latter is always higher because high values are heavily punished by RMSE, and the data contains outliers for every type of metric. For most metrics, the more challenging conditions had a higher difference between MRSE and MJAE, except for the sagittal trunk angle, that remained constant throughout. This means that the trunk angle is the most robust to challenging movements in terms of outliers.

Further analysis of PCEA results before and after applying the offset also shows a marked difference between metrics. The amplitude of the offset showed no impact on the slope of the PCEA line nor its maximum value, meaning that the offset found does not hide any underlying problem with the angles computed. The steepest slopes are found as expected for the metrics having the smallest range of motion: the shank and the trunk.

When comparing PCEA results condition-wise, and grouping them by condition difficulty, we can see that the relative average decrease in all PCEAs (3°, 5° and 15°) is 6.7% between the medium conditions and the easy conditions, and the drop is 15% between the challenging conditions and the easy ones. After the correction, the relative decreases are 19% and 17% respectively, which means that the corrections work best on the easy conditions followed by the challenging conditions and closely by the medium conditions. This can be explained by the cleanliness of the data from easy conditions, with constant offsets but few outliers, meaning that a constant offset correction worked very well with conditions 01 to 04. On the other hand, when the waveforms have a lot of outliers, a constant offset does not help as much.

6.5 Evaluation of the Results According to Literature Guidelines

In section 4.8, we found that the limiting MJAE was limited to 3° for all angles and that any mean over 5° was cause for concern. When analyzing all conditions and both cameras, we found no metrics with a MJAE under 5°, with the closest being the sagittal trunk angle with 6.7°. All other angles had a MJAE of around 10° and 15° for the hips. When only analyzing the L515 for the normal StS though, both the sagittal trunk angle and the shanks were below 5° with respectively 4.4° and 3.4° and the knee angle being at 5.4°, just above the limit.

Correlation coefficients tell a different story, with all CCC from corrected metrics being classified as good or excellent and most being classified as excellent for the easy conditions with the L515. The more difficult conditions have on average a CCC classified as moderate, not sufficient to be used in a clinical setting.

When comparing results with a similar study using the Microsoft Azure Kinect for easy conditions, non-corrected correlation and RMSE for the hip and knee angles are similar for the L515 but worse with the D455 (Thomas et al., 2022). For the corrected results, the RMSE is slightly improved for the L515: 13.9° to our 8.6° for the knee angles and 19.6° to our 15.6° for the hips; the D455 shows worse results. (Thomas et al., 2022).

Comparing to an Opencap system for easy conditions, the corrected data RMSE with the L515 is similar for the knee, but much higher for the hip angle: 7.1° to our 15.6° (Svetek et al., 2025).

6.6 Caveats and Limitations

These results seem absolute, but a few aspects must be considered that can tint the results obtained.

The first aspects to mention are those that worsen the results. Outliers have been discussed extensively in section 5.1.1 and in CHAPTER 6 and are the main problem with the HPE system as it stands. The loss of a single point impacting two metrics and still being considered for all statistics is a factor worsening the results of both the shank angle and the knee angle. The second aspect is the problematic ground-truth trajectory for the femur angle discussed in 5.1.2. This trajectory issue reveals that the ground-truth might not be accurate at all times and that the Vicon system might have been prone to errors during the StS experiment.

The second set of factors discussed here has potentially improved the results of this experiment as compared to the ideal experiment.

The clothes used for the trials were the best possible attire to maximize performance. The skin of the subject was either exposed or covered by a thin adjusted fabric. As demonstrated by conditions 10 and 11 with looser shorts, attire can have a large impact on HPE, so any looser shirt would probably have worsened the hip and trunk metrics even more. Patients who need StS rehabilitation experience mobility impairment which also impairs their ability to put on tight clothes that would be better for the markerless system. As such, this limitation is nontrivial for the population most in need of this technology.

Another aspect related to the target population is that most accurate results were found in conditions 01 to 04 where normal StS repetitions were performed. For most other conditions, an imbalance, asymmetry or exaggeration yielded inferior statistical results. These scenarios are unfortunately bound to happen in clinical settings if this technology is used with patients with mobility impairments, which defeats the purpose.

The offsets applied to the metrics we calculated for all conditions by using as much data as possible to increase generalizability. Unfortunately, only a single subject was in this experiment so we cannot determine if a systematic offset would help all body types for StS. If the problem is related to the calculation of the angles systematically by the HPE then this offset could be beneficial for all, but if it is related to an intrinsic limitation of the HPE having to estimate the joint centres for all body types based on limited information, then the offset could do more harm than good.

6.6.1 Ethical Concerns

The last limitation of this study is related to ethical concerns with AI use in clinical settings and deserves its own section. Machine learning systems often behave in surprising ways, working well for a particular use-case and failing for an adjacent scenario, as we have seen in this study. The training data usually determines what will and will not work, with scenarios seen during training performing much better than unseen ones. Asymmetric StS is a particular scenario that might not be in the training data. It is not an issue for most people using the system, but it ostracizes part of the population. Of course, if a HPE system is built with rehabilitation as a goal, asymmetry and unusual movement related to physical limitation will be included, but this is where the intersectionality of minorities comes in.

Unmasking AI is a 2021 book by Joy Buolamwini talking about facial recognition failures related to skin colour (Buolamwini, 2023). Although this study has not explicitly tested for such biases, the fact that the foot positions were the most problematic and that the shoes worn were black may point to similar issues in visual contrast and detection. In the case of HPE, not only skin tone, but also body types associated with different ethnicities might also be poorly estimated because of their lack of representation in the training data. This becomes especially critical in intersectional cases, when someone belongs to multiple underrepresented groups, such as having a disability and darker skin tone. This person is less likely to receive valid results from a HPE system, while those matching the most common racial group in the region are more likely to benefit from the system.

To prevent technology from reinforcing systemic inequities, it is essential to ensure that AI systems are trained and validated on diverse populations. Fairness must be an explicit design goal, otherwise, the technology risks amplifying the very disparities it aims to solve.

CHAPTER 7 CONCLUSION AND RECOMMENDATIONS

7.1 Summary of Work

This study presented the evaluation of two markerless systems in the context of StS rehabilitation. The main objective was to assess the clinical viability of a low-cost, markerless motion capture system, by comparing its performance against a commercial reference system, with the goal of expanding access to motion-tracking technology for a broader range of patients in need. To that end, two inexpensive commercial 3D cameras, the Intel RealSense D455 and the L515 were tested against a Vicon marker-based system, considered the reference standard in motion analysis. A complete data acquisition and analysis pipeline were developed, integrating Nuitrack SDK for human pose estimation and implementing post-processing steps such as filtering, interpolation, temporal alignment, kinematic computation and statistical evaluation.

The system was validated across 11 conditions designed to simulate a wide range of real-world scenarios, including variations in speed, posture, clothing, and asymmetry. Each condition was recorded with both 3D cameras, and each recording was analyzed across 30 software iterations of the HPE to evaluate the repeatability of the pose estimation output. Performance was evaluated based on both precision (repeatability of the HPE for a given recording) and accuracy (agreement with the Vicon-derived ground truth).

Overall, the L515 outperformed the D455 in both precision and accuracy, in almost all conditions and for almost all joints. In terms of precision, all angles except the shank achieved excellent repeatability, with CCC above 0.9. For the accuracy, most angles achieved a poor agreement prior to the offset correction approach. This correction applied a systematic bias based on the middle 50% of all residuals per joint type. This improved key metrics across the board: the average MJAE dropped from over 23° to 11.2°, the RMSE fell to 15.6°, and PCEA@5 increased from 5.8% to 46%. While the shank angles showed poor CCC performance, the knee and the femur angle achieved a better, moderate agreement based on CCC. Although still not within strict clinical thresholds, these improvements suggest that systematic error correction can significantly enhance the performance of markerless systems.

Looking back on the hypotheses stated at the beginning of this document:

- Hypothesis #1 was that a single device plug-and-play 3D camera system would be convenient for StS evaluation, which proved correct, proven by the successful and simple experiment.
- Hypothesis #2 stated that StS could be accurately evaluated for standard conditions, but this proved only partially true since exclusively the shank, knee and trunk angle having an acceptable MJAE after correction.
- Hypothesis #3 stated that challenging conditions could also be evaluated with the system, but this proved false as both cameras showed sensitivity to challenging conditions.

7.2 Contributions and Implications

This work provides a proof of concept that shows the potential of markerless human pose estimation in rehabilitation. It provides a methodology for integrating low-cost 3D hardware with freely available pose estimation software and systematically evaluates their performance against gold-standard motion capture data. The findings demonstrate that while raw output from such systems may not yet meet clinical standards for all use cases, they can yield clinically meaningful data for specific tasks and under well-controlled conditions, particularly if systematic biases are addressed.

Moreover, this thesis emphasizes the importance of evaluating not just global accuracy but also performance under realistic, real-world condition. The inclusion of challenges such as loose clothing, asymmetrical movements, and nonstandard postures was essential to stress-test the system in scenarios that more closely resemble clinical settings.

This work also highlights ethical and demographic considerations in the deployment of machine learning based pose estimation. The accuracy and robustness of these systems can vary depending on the population with variation in clothing, body type, and movement characteristics. Ensuring equitable performance across diverse patient populations requires the intentional creation of specific datasets for training, something essential if such technologies are to be trusted in healthcare.

7.3 Future Research

Future work should focus on training models on a higher variety of movement pattern, asymmetry and body types in order to extend the performance obtained to more complex scenarios, likely to appear in clinical settings. Additionally, state-of-the-art off-the-shelf 3D camera systems and HPE systems should continue to be evaluated on a regular basis, as their current performance is nearing clinical viability and the underlying technology is evolving rapidly. Establishing standardized benchmarks and testing protocols would also help guide future improvements and ensure these systems can be reliably adopted in rehabilitation settings.

REFERENCES

Akhtaruzzaman, M. D., Shafie, A. A., & Khan, M. R. (2016). GAIT ANALYSIS: SYSTEMS, TECHNOLOGIES, AND IMPORTANCE. *Journal of Mechanics in Medicine and Biology*, *16*(07), 1630003. https://doi.org/10.1142/s0219519416300039

Avogaro, A., Cunico, F., Rosenhahn, B., & Setti, F. (2023). Markerless human pose estimation for biomedical applications: a survey. *Frontiers in Computer Science*, *5*, 1153160. https://doi.org/10.3389/fcomp.2023.1153160

Bertec. (2025). *Force Plates*. https://www.bertec.com/products/force-plates

Bhardwaj, S., Khan, A. A., & Muzammil, M. (2021). Lower limb rehabilitation using multimodal measurement of sit-to-stand and stand-to-sit task. *Disability and Rehabilitation: Assistive Technology, 16*(5), 438-445. https://doi.org/10.1080/17483107.2019.1629701

BiometricsLTD. (2024). *Electronic Goniometers*. https://www.biometricsltd.com/goniometer.htm

Boswell, M. A., Kidziński, Ł., Hicks, J. L., Uhlrich, S. D., Falisse, A., & Delp, S. L. (2023). Smartphone videos of the sit-to-stand test predict osteoarthritis and health outcomes in a nationwide study. *npj Digital Medicine*, *6*(1), 32. https://doi.org/10.1038/s41746-023-00775-1

Buolamwini, J. (2023). *Unmasking AI: my mission to protect what is human in a world of machines*. Random House.

Campitelli, A., Paulson, S., Vincenzo, J., Glenn, J. M., Gills, J. L., Jones, M. D., Powers, M., & Gray, M. (2021). Sit-to-Stand Power Across the Lifespan: A Cross-Sectional Analysis. *Journal of aging and physical activity*, *30*(4), 678-688. https://doi.org/10.1123/japa.2021-0066

Cao, Z., Hidalgo, G., Simon, T., Wei, S.-E., & Sheikh, Y. (2018). OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields. *arXiv*. https://doi.org/10.48550/arxiv.1812.08008

Cardona, M., Solanki, V. K., & Cena, C. E. G. (2020). Exoskeleton Robots for Rehabilitation and Healthcare Devices. *SpringerBriefs in Applied Sciences and Technology*. https://doi.org/10.1007/978-981-15-4732-4

Chaovalit, S., Taylor, N. F., & Dodd, K. J. (2020). Sit-to-stand exercise programs improve sit-to-stand performance in people with physical impairments due to health conditions: a systematic review and meta-analysis. *Disabil Rehabil*, *42*(9), 1202-1211. https://doi.org/10.1080/09638288.2018.1524518

Cherry-Allen, K. M., French, M. A., Stenum, J., Xu, J., & Roemmich, R. T. (2023). Opportunities for Improving Motor Assessment and Rehabilitation After Stroke by Leveraging Video-Based Pose Estimation. *American Journal of Physical Medicine & Rehabilitation*, *102*(2S), S68-S74. https://doi.org/10.1097/phm.0000000000002131

Clark, R. A., Mentiplay, B. F., Hough, E., & Pua, Y. H. (2019). Three-dimensional cameras and skeleton pose tracking for physical function assessment: A review of uses, validity, current developments and Kinect alternatives. *Gait & Posture*, *68*, 193-200. https://doi.org/10.1016/j.gaitpost.2018.11.029

Cong, P., Xu, Y., Ren, Y., Zhang, J., Xu, L., Wang, J., Yu, J., & Ma, Y. (2022). Weakly Supervised 3D Multi-person Pose Estimation for Large-scale Scenes based on Monocular Camera and Single LiDAR. *arXiv*. https://doi.org/10.48550/arxiv.2211.16951

Dolatabadi, E., Taati, B., & Mihailidis, A. (2016). Automated classification of pathological gait after stroke using ubiquitous sensing technology. *Annu Int Conf IEEE Eng Med Biol Soc*, *2016*, 6150-6153. https://doi.org/10.1109/embc.2016.7592132

Efstratiou, & Panagiotis. (2021). SKELETON- TRACKING FOR SPORTS USING LiDAR DEPTH CAMERA. https://app.readcube.com/library/0b762f35-1fb5-439f-a5dd-1c9b528858e5/item/2f1b09af-6642-46c8-a3b4-a4336ac0f20d

Ekberg, P., Daemi, B., & Mattsson, L. (2017). 3D precision measurements of meter-sized surfaces using low cost illumination and camera techniques. *Measurement Science and Technology*, *28*. https://doi.org/10.1088/1361-6501/aa5ae6

Feldhege, F., Mau-Moeller, A., Lindner, T., Hein, A., Markschies, A., Zettl, U. K., & Bader, R. (2015). Accuracy of a Custom Physical Activity and Knee Angle Measurement Sensor System for Patients with Neuromuscular Disorders and Gait Abnormalities. *Sensors*, *15*(5), 10734-10752. https://www.mdpi.com/1424-8220/15/5/10734

Fernández-Gorgojo, M., Salas-Gómez, D., Sánchez-Juan, P., Barbado, D., Laguna-Bercero, E., & Pérez-Núñez, M. I. (2022). Clinical–Functional Evaluation and Test–Retest Reliability of the G-WALK Sensor in Subjects with Bimalleolar Ankle Fractures 6 Months after Surgery. *Sensors*, *22*(8), 3050. https://doi.org/10.3390/s22083050

Five Times Sit to Stand (5TSTS). (2019). In A. o. N. P. Therapy (Ed.).

Fürst, M., Gupta, S. T. P., Schuster, R., Wasenmüller, O., & Stricker, D. (2020). HPERL: 3D Human Pose Estimation from RGB and LiDAR. *arXiv*. https://doi.org/10.48550/arxiv.2010.08221

Glandon, A., Vidyaratne, L., Sadeghzadehyazdi, N., Dhar, N. K., Familoni, J. O., Acton, S. T., & Iftekharuddin, K. M. (2019, 14-19 July 2019). 3D Skeleton Estimation and Human Identity Recognition Using Lidar Full Motion Video. 2019 International Joint Conference on Neural Networks (IJCNN),

GovernmentOfCanada. (2023). *Stroke in Canada*. https://www.canada.ca/en/public-health/services/publications/diseases-conditions/stroke-in-canada.html

Guan, S., Gray, H. A., Keynejad, F., & Pandy, M. G. (2016). Mobile Biplane X-Ray Imaging System for Measuring 3D Dynamic Joint Motion During Overground Gait. *IEEE Transactions on Medical Imaging*, *35*(1), 326-336. https://doi.org/10.1109/tmi.2015.2473168

Hogg, R. V., Tanis, E. A., & Zimmerman, D. L. (2020). *Probability and Statistical Inference*. Pearson. https://books.google.ca/books?id=xc6ltgEACAAJ

Horsak, B., Prock, K., Krondorfer, P., Siragy, T., Simonlehner, M., & Dumphart, B. (2024). Inter-trial variability is higher in 3D markerless compared to marker-based motion capture: Implications for data post-processing and analysis. *Journal of Biomechanics*, *166*, 112049. https://doi.org/10.1016/j.jbiomech.2024.112049

Hullfish, T. J., Qu, F., Stoeckl, B. D., Gebhard, P. M., Mauck, R. L., & Baxter, J. R. (2019). Measuring clinically relevant knee motion with a self-calibrated wearable sensor. *Journal of Biomechanics*, *89*, 105-109. https://doi.org/https://doi.org/10.1016/j.jbiomech.2019.04.003

Hyodo, K., Kanamori, A., Kadone, H., Kajiwara, M., Okuno, K., Kikuchi, N., & Yamazaki, M. (2023). Evaluation of Sit-to-Stand Movement Focusing on Kinematics, Kinetics, and Muscle Activity after Modern Total Knee Arthroplasty. *Applied Sciences*, *14*(1), 360. https://doi.org/10.3390/app14010360

IntelRealSense. (2025a). *Intel RealSense Depth Camera D455*. Retrieved 2025-05-27 from https://www.intelrealsense.com/depth-camera-d455/

IntelRealSense. (2025b). *Intel RealSense Lidar Camera L515*. Retrieved 2025-05-27 from https://www.intelrealsense.com/lidar-camera-l515/

Janssen, W. G. M., Bussmann, H. B. J., & Stam, H. J. (2002). Determinants of the Sit-to-Stand Movement: A Review. *Physical Therapy*, *82*(9), 866-879. https://doi.org/10.1093/ptj/82.9.866

Kaewrat, C., Khundam, C., & Thu, M. (2024). Enhancing Exercise Monitoring and Guidance Through Mobile Augmented Reality: A Comparative Study of RGB and LiDAR. *IEEE Access, 12*, 95447-95460. https://doi.org/10.1109/access.2024.3425163

Kasven-Gonzalez, N., Souverain, R., & Miale, S. (2010). Improving quality of life through rehabilitation in palliative care: Case report. *Palliative and Supportive Care*, *8*(3), 359-369. https://doi.org/10.1017/S1478951510000167

Kiatkulanusorn, S., Luangpon, N., Srijunto, W., Watechagit, S., Pitchayadejanant, K., Kuharat, S., Bég, O. A., & Suato, B. P. (2023). Analysis of the concurrent validity and reliability of five common clinical goniometric devices. *Scientific Reports*, *13*(1), 20931. https://doi.org/10.1038/s41598-023-48344-6

Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J Chiropr Med*, *15*(2), 155-163. https://doi.org/10.1016/j.jcm.2016.02.012

Kumar, T., Rao, R. R., Ramamurthy, P. C., & Mani, M. (2018). Safety of Light Emitting Diode (LED) Based Domestic Lighting in Rural Context. *2018 15th IEEE India Council International Conference (INDICON), 00*, 1-5. https://doi.org/10.1109/indicon45594.2018.8987093

Lin, T.-Y., Maire, M., Belongie, S., Bourdev, L., Girshick, R., Hays, J., Perona, P., Ramanan, D., Zitnick, C. L., & Dollár, P. (2014). Microsoft COCO: Common Objects in Context. *arXiv*. https://doi.org/10.48550/arxiv.1405.0312

Mahmood, I., Maqbool, H. F., Raza, A., Iqbal, N., & Dehghani-Sanij, A. A. (2024). Gait dynamic stability evaluation in patients undergoing hip joint fractures – tools to measure rehabilitation effectiveness. *Biomedical Physics & Engineering Express*, *10*(4), 045050. https://doi.org/10.1088/2057-1976/ad567b

Mao, Y. R., Wu, X. Q., Zhao, J. L., Lo, W. L. A., Chen, L., Ding, M. H., Xu, Z. Q., Bian, R. H., Huang, D. F., & Li, L. (2018). The Crucial Changes of Sit-to-Stand Phases in Subacute

Stroke Survivors Identified by Movement Decomposition Analysis. *Frontiers in Neurology*, *9*, 185. https://doi.org/10.3389/fneur.2018.00185

Martınez, G. H. (2019). Openpose: Whole-body pose estimation. *Ph. D. dissertation*.

Masullo, A., Burghardt, T., Perrett, T., Damen, D., & Mirmehdi, M. (2019). Sit-to-Stand Analysis in the Wild using Silhouettes for Longitudinal Health Monitoring. *arXiv*. https://doi.org/10.48550/arxiv.1910.01370

McGinley, J. L., Baker, R., Wolfe, R., & Morris, M. E. (2009). The reliability of three-dimensional kinematic gait measurements: a systematic review. *Gait & Posture*, *29*(3), 360-369. https://doi.org/10.1016/j.gaitpost.2008.09.003

Movella. *Xsens Awinda Starter*. https://www.movella.com/products/motion-capture/xsens-mvn-awinda-starter?_gl=1*14lgzh9*_up*MQ..*_gs*MQ..&gclid=Cj0KCQjw097CBhDIARIsAJ3-nxdFaqyRXHuQ1HlaMp-MOWod7usVdcqY21Bv4G488uBgm1sul757IUcaArPgEALw_wcB&gbraid=0AAAAAD_wMeJ_7wvzdnIsrwQquMVg1FKzR

Mündermann, L., Corazza, S., & Andriacchi, T. P. (2006). The evolution of methods for the capture of human movement leading to markerless motion capture for biomechanical applications. *Journal of NeuroEngineering and Rehabilitation*, *3*(1), 6. https://doi.org/10.1186/1743-0003-3-6

NuitrackSDK. (2024). *Nuitrack Blog*. https://nuitrack.com/blog

Riazati, S., McGuirk, T. E., Perry, E. S., Sihanath, W. B., & Patten, C. (2022). Absolute Reliability of Gait Parameters Acquired With Markerless Motion Capture in Living Domains [Original Research]. *Frontiers in Human Neuroscience*, *Volume 16 - 2022*. https://doi.org/10.3389/fnhum.2022.867474

Scataglini, S., Abts, E., Bocxlaer, C. V., Bussche, M. V. d., Meletani, S., & Truijen, S. (2024). Accuracy, Validity, and Reliability of Markerless Camera-Based 3D Motion Capture Systems versus Marker-Based 3D Motion Capture Systems in Gait Analysis: A Systematic Review and Meta-Analysis. *Sensors*, *24*(11), 3686. https://doi.org/10.3390/s24113686

Schenkman, M., Berger, R. A., Riley, P. O., Mann, R. W., & Hodge, W. A. (1990). Whole-body movements during rising to standing from sitting. *Phys Ther*, *70*(10), 638-648; discussion 648-651. https://doi.org/10.1093/ptj/70.10.638

Selva, J., Johansen, A. S., Escalera, S., Nasrollahi, K., Moeslund, T. B., & Clapés, A. (2022). Video Transformers: A Survey. *arXiv*. https://doi.org/10.48550/arxiv.2201.05991

Serrancolí, G., Bogatikov, P., Huix, J. P., Barberà, A. F., Egea, A. J. S., Ribé, J. T., Kanaan-Izquierdo, S., & Susín, A. (2020). Marker-Less Monitoring Protocol to Analyze Biomechanical Joint Metrics During Pedaling. *IEEE Access*, *8*, 122782-122790. https://doi.org/10.1109/ACCESS.2020.3006423

Seth, A., Hicks, J. L., Uchida, T. K., Habib, A., Dembia, C. L., Dunne, J. J., Ong, C. F., DeMers, M. S., Rajagopal, A., Millard, M., Hamner, S. R., Arnold, E. M., Yong, J. R., Lakshmikanth, S. K., Sherman, M. A., Ku, J. P., & Delp, S. L. (2018). OpenSim: Simulating musculoskeletal dynamics and neuromuscular control to study human and animal movement. *PLoS Comput Biol*, *14*(7), e1006223. https://doi.org/10.1371/journal.pcbi.1006223

Stenum, J., Rossi, C., & Roemmich, R. T. (2021). Two-dimensional video-based analysis of human gait using pose estimation. *PLoS Comput Biol*, *17*(4), e1008935. https://doi.org/10.1371/journal.pcbi.1008935

Svetek, A., Morgan, K., Burland, J., & Glaviano, N. R. (2025). Validation of OpenCap on lower extremity kinematics during functional tasks. *Journal of Biomechanics*, *183*, 112602. https://doi.org/https://doi.org/10.1016/j.jbiomech.2025.112602

TheiaMarkerless. (2025). https://www.theiamarkerless.com/

Thomas, J., Hall, J. B., Bliss, R., & Guess, T. M. (2022). Comparison of Azure Kinect and optical retroreflective motion capture for kinematic and spatiotemporal evaluation of the sit-to-stand test. *Gait & Posture*, *94*, 153-159. https://doi.org/https://doi.org/10.1016/j.gaitpost.2022.03.011

Tramontano, M., Belluscio, V., Bergamini, E., Allevi, G., Angelis, S. D., Verdecchia, G., Formisano, R., Vannozzi, G., & Buzzi, M. G. (2022). Vestibular Rehabilitation Improves Gait Quality and Activities of Daily Living in People with Severe Traumatic Brain Injury: A Randomized Clinical Trial. *Sensors*, *22*(21), 8553. https://doi.org/10.3390/s22218553

Uhlrich, S. D., Falisse, A., Kidziński, Ł., Muccini, J., Ko, M., Chaudhari, A. S., Hicks, J. L., & Delp, S. L. (2023). OpenCap: Human movement dynamics from smartphone videos. *PLOS Computational Biology*, *19*(10), e1011462. https://doi.org/10.1371/journal.pcbi.1011462

Vicon. (2025a). *Full body modeling with Plug-in Gait*. https://help.vicon.com/space/Nexus216/11607226/Full+body+modeling+with+Plug-in+Gait

Vicon. (2025b). *Valkyrie*. https://www.vicon.com/hardware/cameras/valkyrie/

Wang, H., Xu, S., Fu, J., Xu, X., Wang, Z., & Na, R. S. (2022). Sit-to-Stand (STS) Movement Analysis of the Center of Gravity for Human–Robot Interaction. *Frontiers in Neurorobotics*, *16*, 863722. https://doi.org/10.3389/fnbot.2022.863722

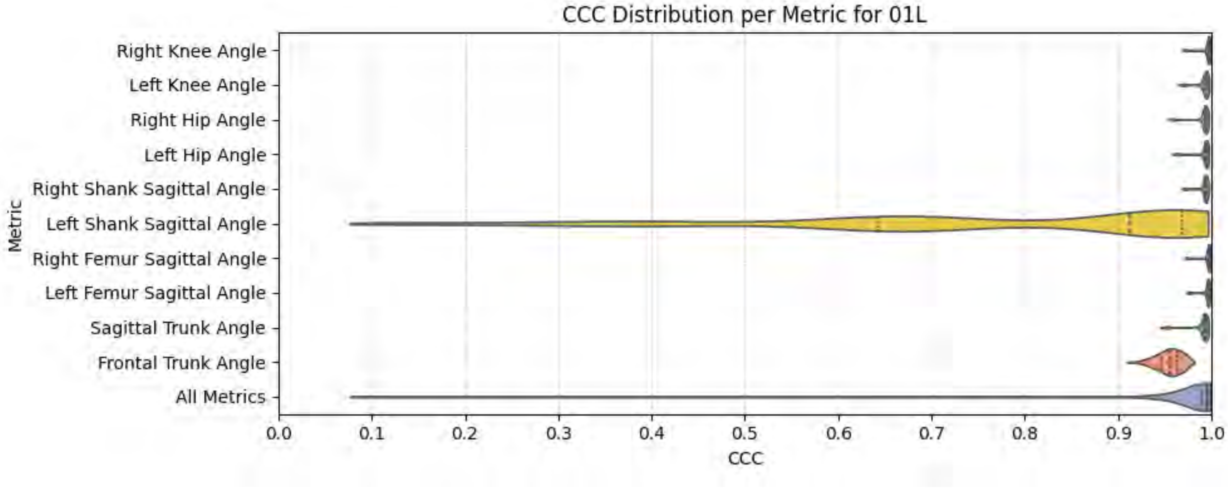
Washabaugh, E. P., Shanmugam, T. A., Ranganathan, R., & Krishnan, C. (2022). Comparing the accuracy of open-source pose estimation methods for measuring gait kinematics. *Gait & Posture*, *97*, 188-195. https://doi.org/10.1016/j.gaitpost.2022.08.008

Xu, T., An, D., Jia, Y., & Yue, Y. (2021). A Review: Point Cloud-Based 3D Human Joints Estimation. *Sensors*, *21*(5), 1684. https://doi.org/10.3390/s21051684

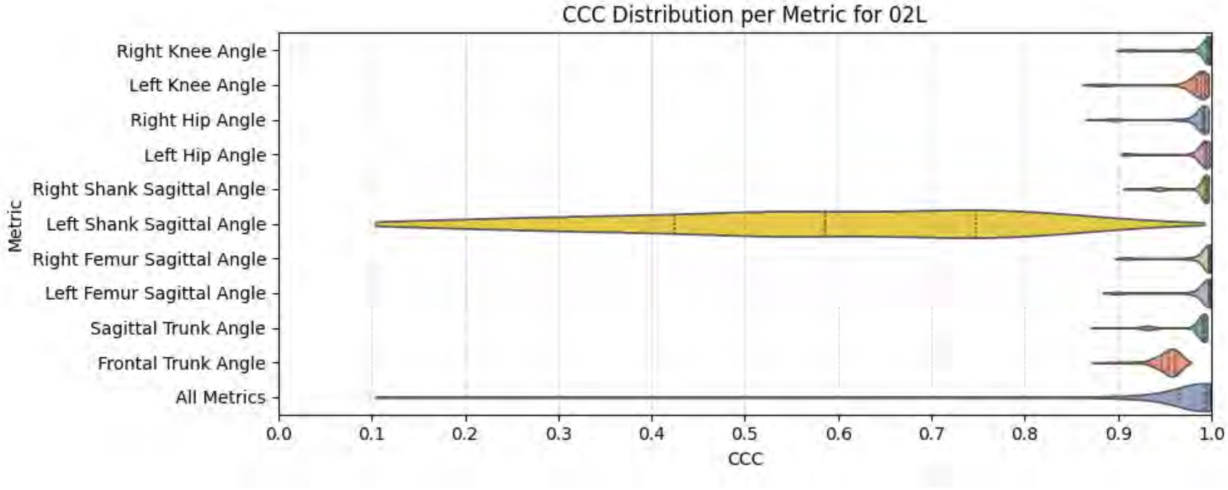
Yang, F., Zhong, Z., Luo, Z., Cai, Y., Lin, Y., Li, S., & Sebe, N. (2021). Joint Noise-Tolerant Learning and Meta Camera Shift Adaptation for Unsupervised Person Re-Identification. *arXiv*. https://doi.org/10.48550/arxiv.2103.04618

Zlatičanin, R., Jaganjac, A., & Kraljević, B. (2024). APPLICATION OF DIFFERENT EXERCISE PROGRAMS IN THE PREVENTION OF FALLS IN PEOPLE OF THE THIRD AGE. *Homosporticus*, 48-54. https://doi.org/10.61886/1840-4324.2024.26.1.48

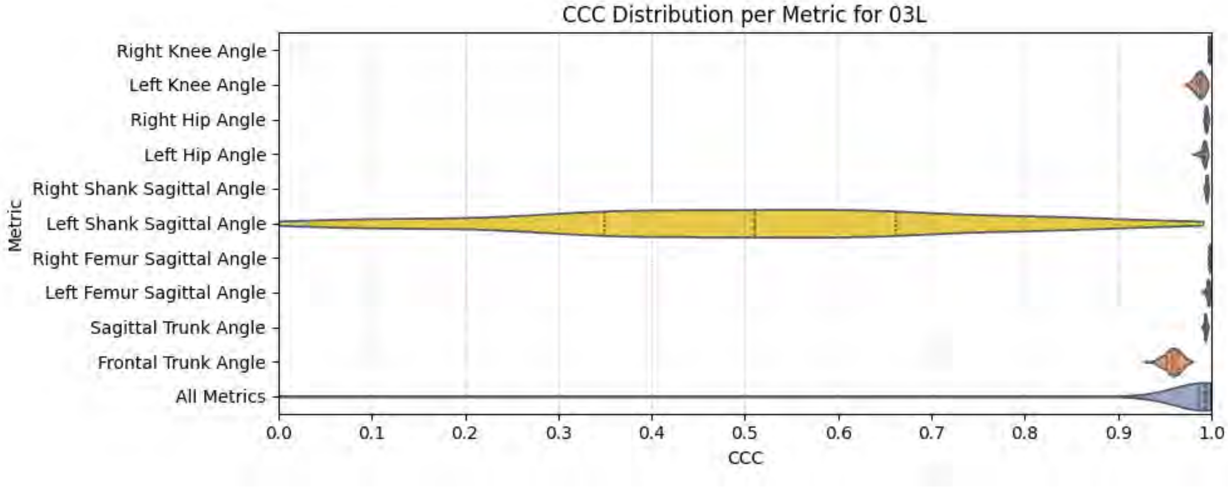
APPENDIX A COMPLETE PRECISION RESULTS



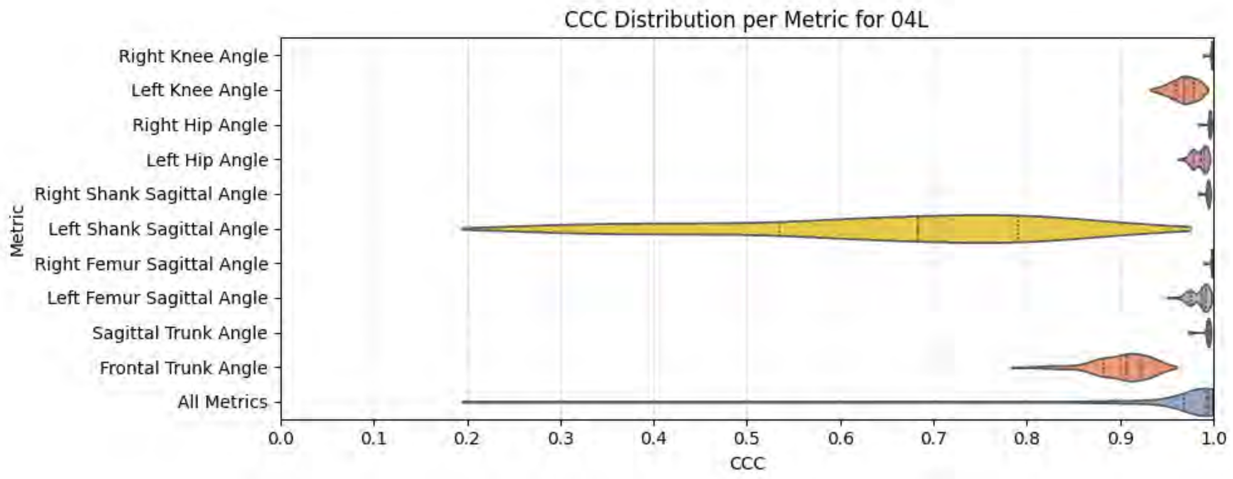
a)



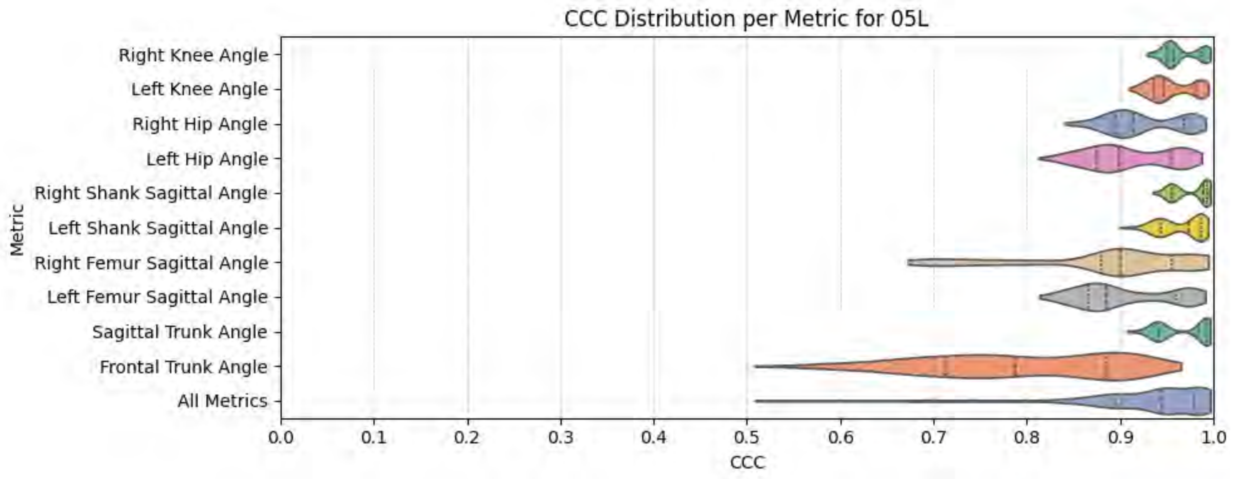
b)



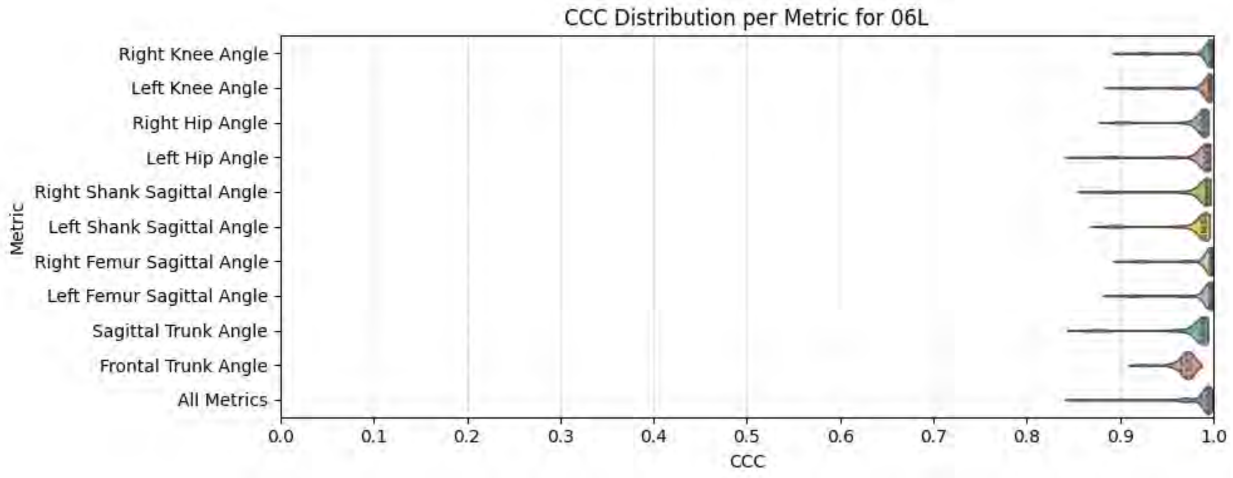
c)


CCC Distribution per Metric for 04L
Metric
Right Knee Angle
Left Knee Angle
Right Hip Angle
Left Hip Angle
Right Shank Sagittal Angle
Left Shank Sagittal Angle
Right Femur Sagittal Angle
Left Femur Sagittal Angle
Sagittal Trunk Angle
Frontal Trunk Angle
All Metrics
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
CCC

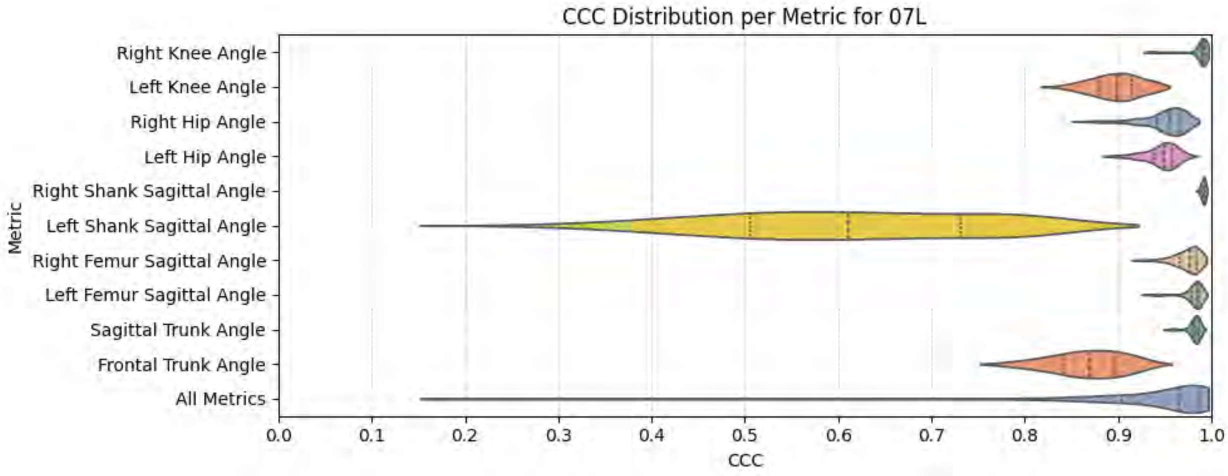
d)


CCC Distribution per Metric for 05L
Metric
Right Knee Angle
Left Knee Angle
Right Hip Angle
Left Hip Angle
Right Shank Sagittal Angle
Left Shank Sagittal Angle
Right Femur Sagittal Angle
Left Femur Sagittal Angle
Sagittal Trunk Angle
Frontal Trunk Angle
All Metrics
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
CCC

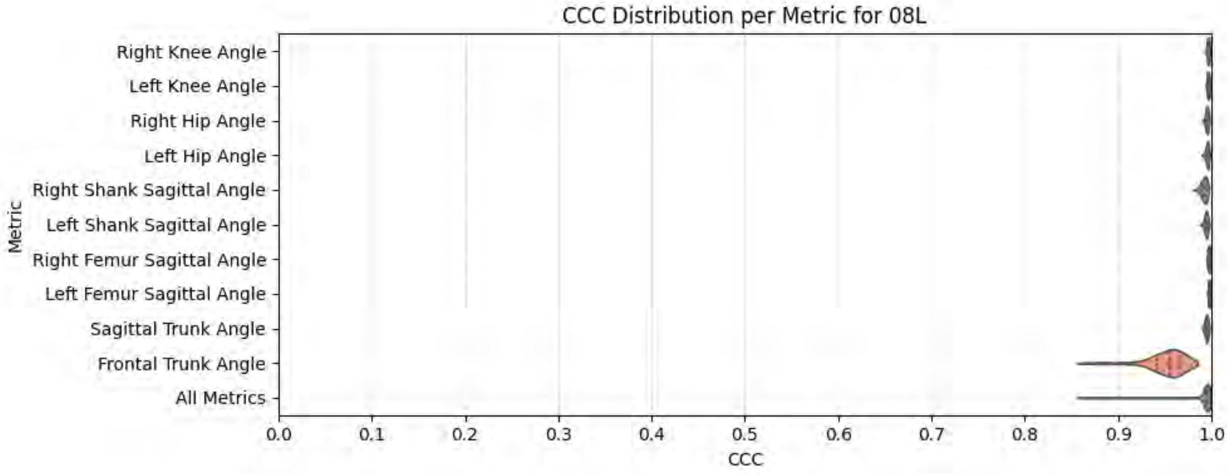
e)


CCC Distribution per Metric for 06L
Metric
Right Knee Angle
Left Knee Angle
Right Hip Angle
Left Hip Angle
Right Shank Sagittal Angle
Left Shank Sagittal Angle
Right Femur Sagittal Angle
Left Femur Sagittal Angle
Sagittal Trunk Angle
Frontal Trunk Angle
All Metrics
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
CCC

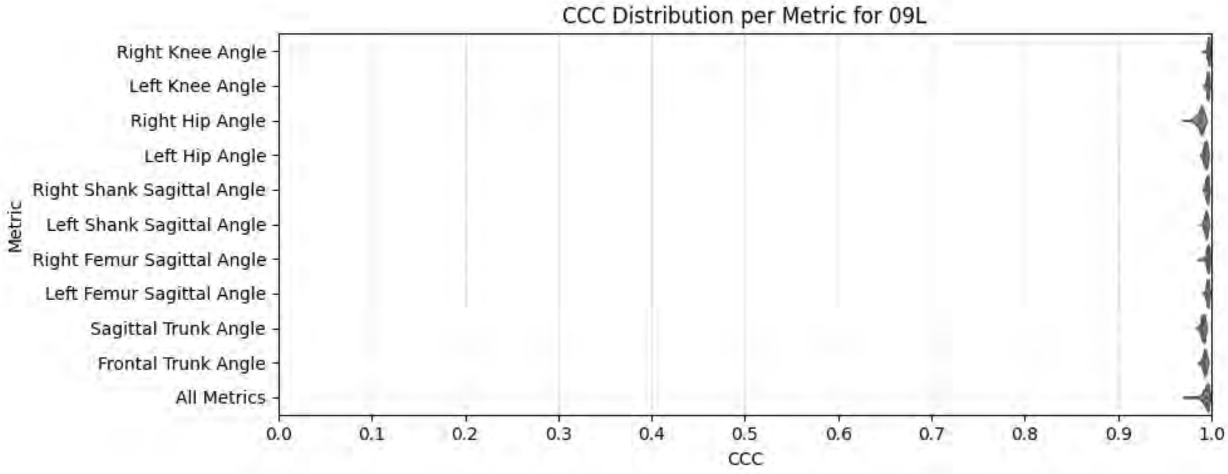
f)


CCC Distribution per Metric for 07L
Right Knee Angle
Left Knee Angle
Right Hip Angle
Left Hip Angle
Right Shank Sagittal Angle
Left Shank Sagittal Angle
Right Femur Sagittal Angle
Left Femur Sagittal Angle
Sagittal Trunk Angle
Frontal Trunk Angle
All Metrics
Metric
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
CCC

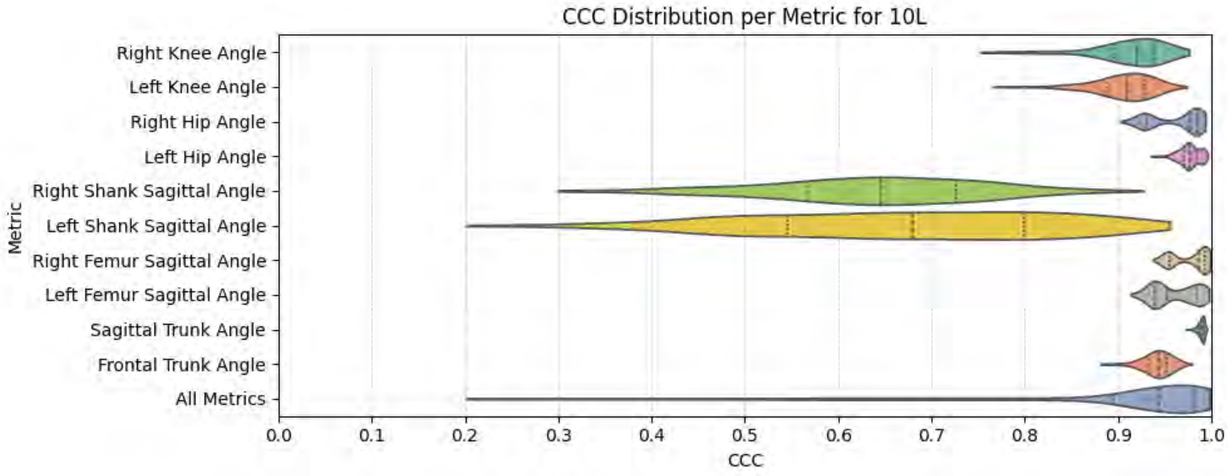
g)


CCC Distribution per Metric for 08L
Right Knee Angle
Left Knee Angle
Right Hip Angle
Left Hip Angle
Right Shank Sagittal Angle
Left Shank Sagittal Angle
Right Femur Sagittal Angle
Left Femur Sagittal Angle
Sagittal Trunk Angle
Frontal Trunk Angle
All Metrics
Metric
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
CCC

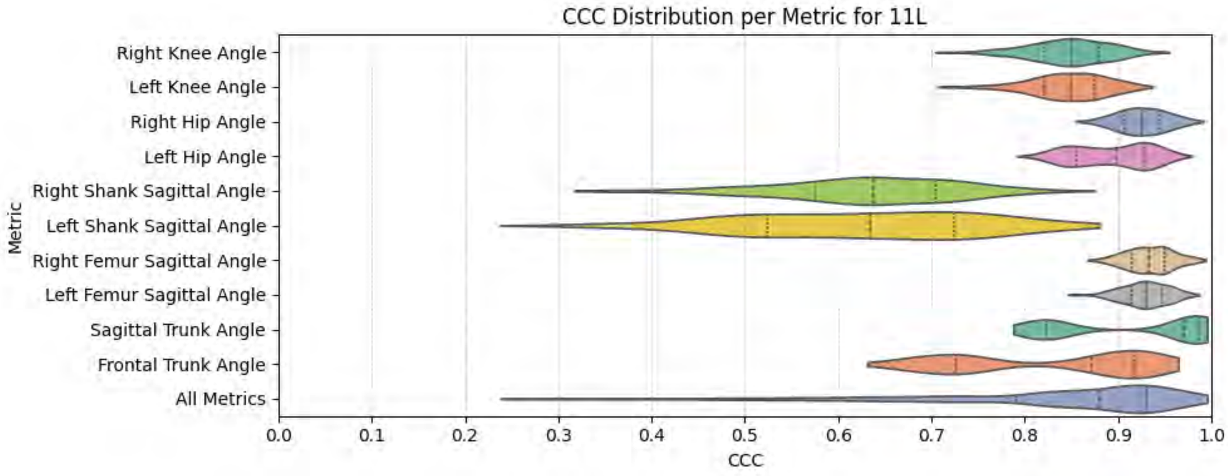
h)


CCC Distribution per Metric for 09L
Right Knee Angle
Left Knee Angle
Right Hip Angle
Left Hip Angle
Right Shank Sagittal Angle
Left Shank Sagittal Angle
Right Femur Sagittal Angle
Left Femur Sagittal Angle
Sagittal Trunk Angle
Frontal Trunk Angle
All Metrics
Metric
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
CCC

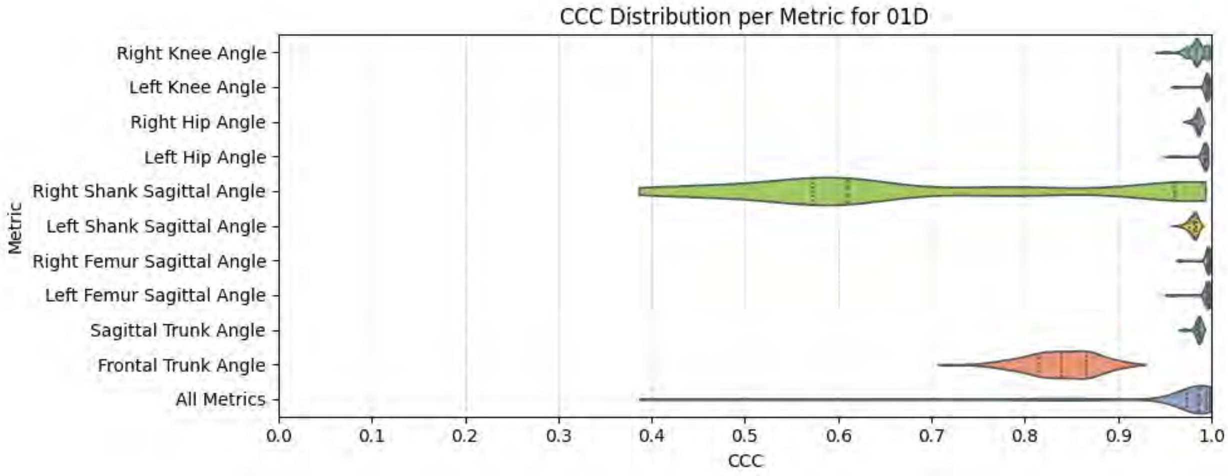
i)


CCC Distribution per Metric for 10L
Metric
Right Knee Angle
Left Knee Angle
Right Hip Angle
Left Hip Angle
Right Shank Sagittal Angle
Left Shank Sagittal Angle
Right Femur Sagittal Angle
Left Femur Sagittal Angle
Sagittal Trunk Angle
Frontal Trunk Angle
All Metrics
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
CCC

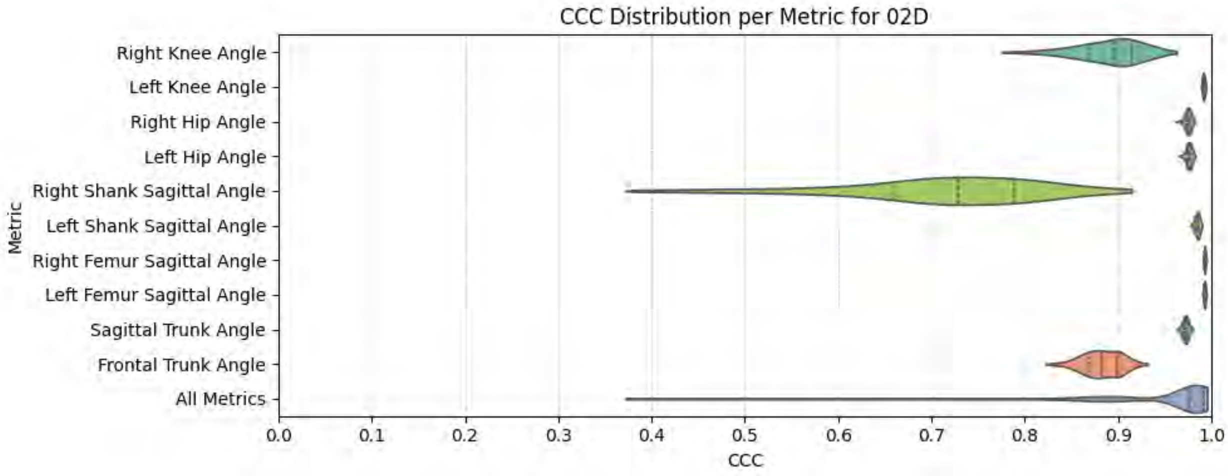
j)


CCC Distribution per Metric for 11L
Metric
Right Knee Angle
Left Knee Angle
Right Hip Angle
Left Hip Angle
Right Shank Sagittal Angle
Left Shank Sagittal Angle
Right Femur Sagittal Angle
Left Femur Sagittal Angle
Sagittal Trunk Angle
Frontal Trunk Angle
All Metrics
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
CCC

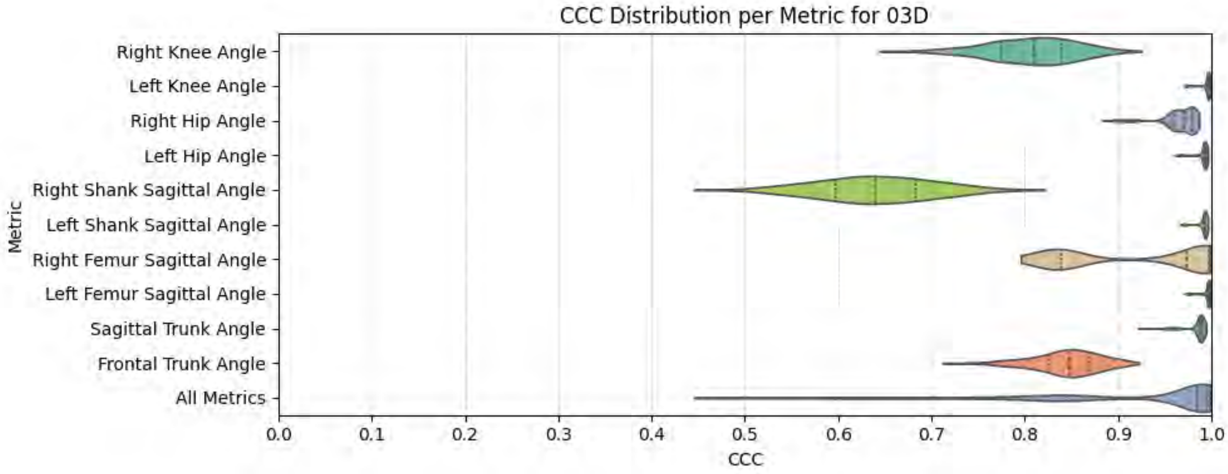
k)


CCC Distribution per Metric for 01D
Metric
Right Knee Angle
Left Knee Angle
Right Hip Angle
Left Hip Angle
Right Shank Sagittal Angle
Left Shank Sagittal Angle
Right Femur Sagittal Angle
Left Femur Sagittal Angle
Sagittal Trunk Angle
Frontal Trunk Angle
All Metrics
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
CCC

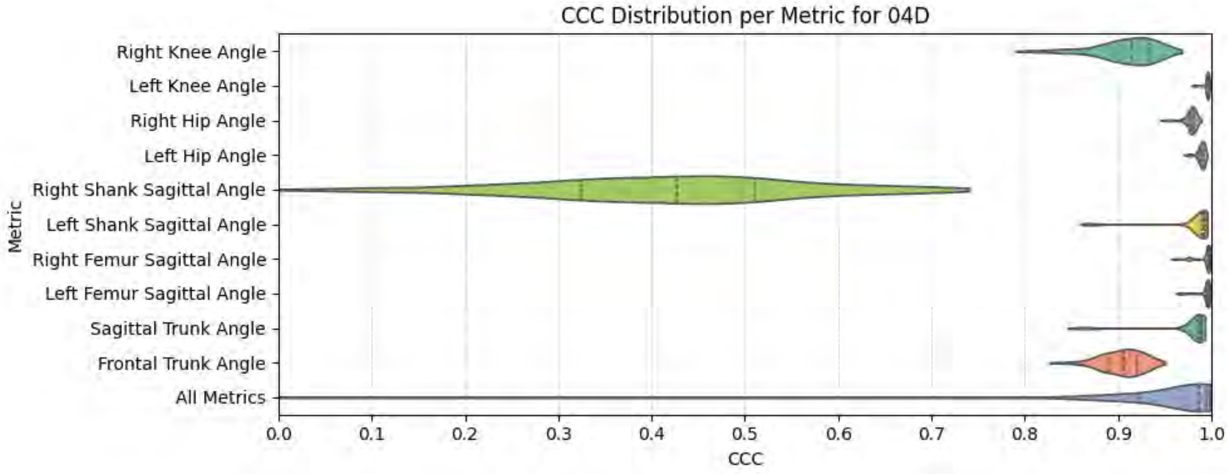
l)


CCC Distribution per Metric for 02D
Metric
Right Knee Angle
Left Knee Angle
Right Hip Angle
Left Hip Angle
Right Shank Sagittal Angle
Left Shank Sagittal Angle
Right Femur Sagittal Angle
Left Femur Sagittal Angle
Sagittal Trunk Angle
Frontal Trunk Angle
All Metrics
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
CCC

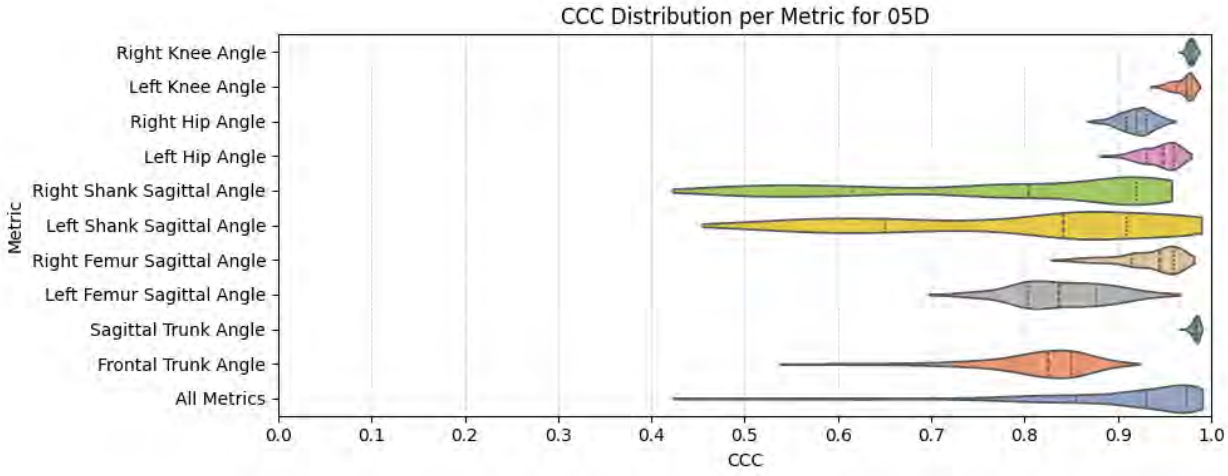
m)


CCC Distribution per Metric for 03D
Metric
Right Knee Angle
Left Knee Angle
Right Hip Angle
Left Hip Angle
Right Shank Sagittal Angle
Left Shank Sagittal Angle
Right Femur Sagittal Angle
Left Femur Sagittal Angle
Sagittal Trunk Angle
Frontal Trunk Angle
All Metrics
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
CCC

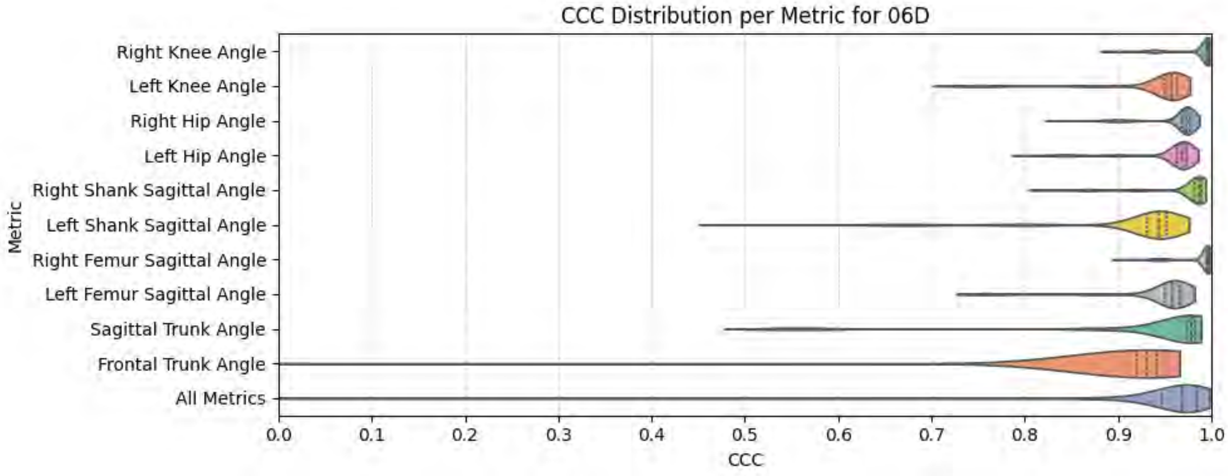
n)


CCC Distribution per Metric for 04D
Metric
Right Knee Angle
Left Knee Angle
Right Hip Angle
Left Hip Angle
Right Shank Sagittal Angle
Left Shank Sagittal Angle
Right Femur Sagittal Angle
Left Femur Sagittal Angle
Sagittal Trunk Angle
Frontal Trunk Angle
All Metrics
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
CCC

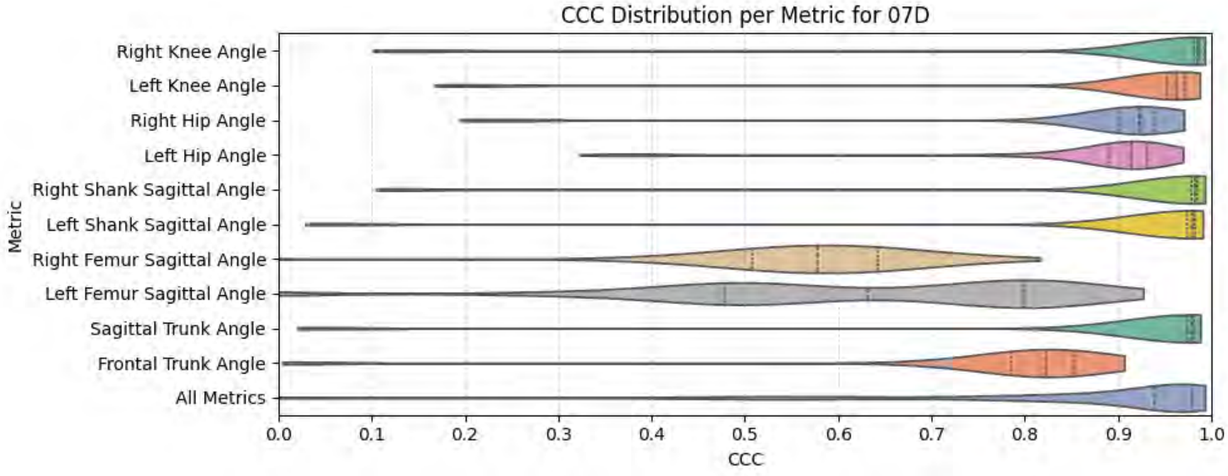
o)


CCC Distribution per Metric for 05D
Metric
Right Knee Angle
Left Knee Angle
Right Hip Angle
Left Hip Angle
Right Shank Sagittal Angle
Left Shank Sagittal Angle
Right Femur Sagittal Angle
Left Femur Sagittal Angle
Sagittal Trunk Angle
Frontal Trunk Angle
All Metrics
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
CCC

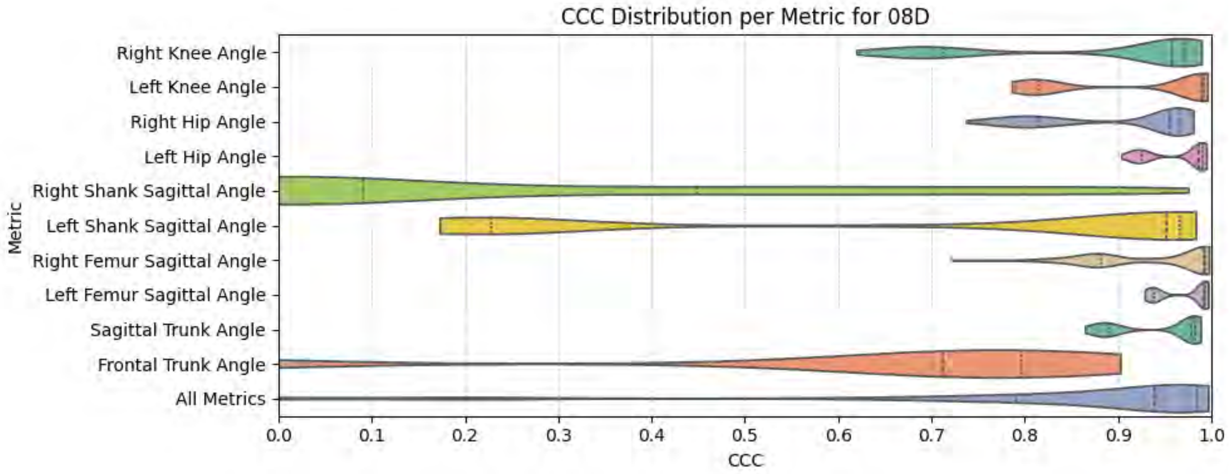
p)


CCC Distribution per Metric for 06D
Metric
Right Knee Angle
Left Knee Angle
Right Hip Angle
Left Hip Angle
Right Shank Sagittal Angle
Left Shank Sagittal Angle
Right Femur Sagittal Angle
Left Femur Sagittal Angle
Sagittal Trunk Angle
Frontal Trunk Angle
All Metrics
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
CCC

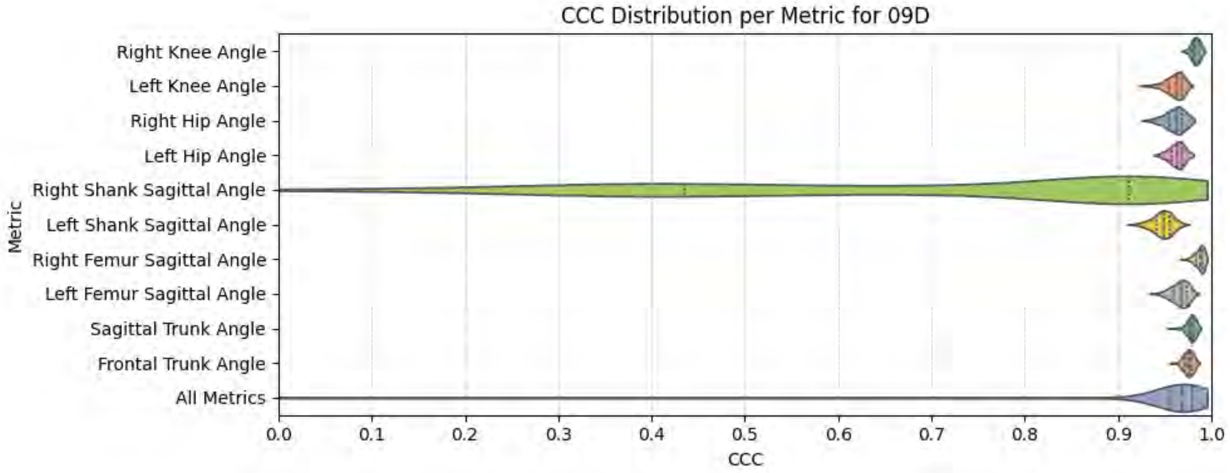
q)


CCC Distribution per Metric for 07D
Metric
Right Knee Angle
Left Knee Angle
Right Hip Angle
Left Hip Angle
Right Shank Sagittal Angle
Left Shank Sagittal Angle
Right Femur Sagittal Angle
Left Femur Sagittal Angle
Sagittal Trunk Angle
Frontal Trunk Angle
All Metrics
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
CCC

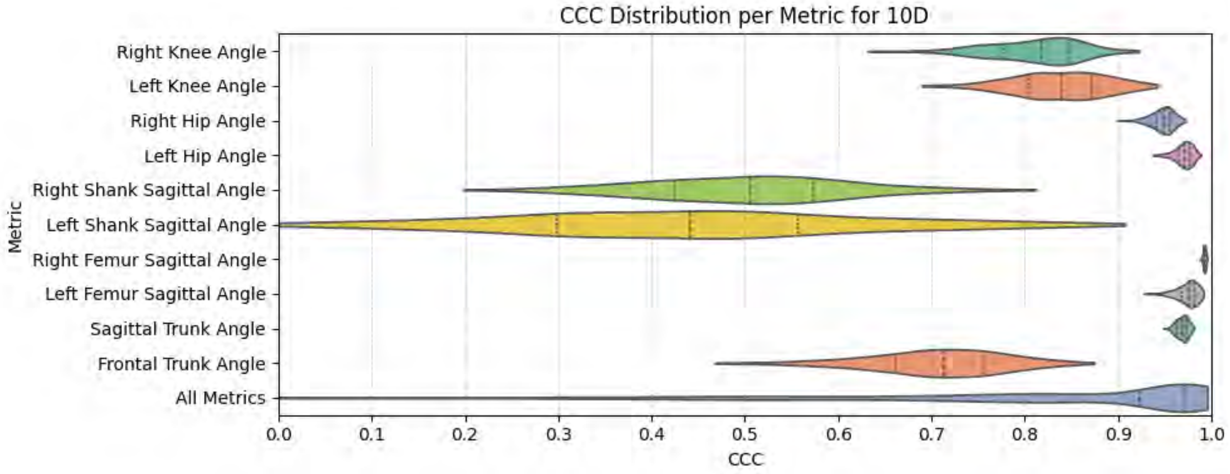
r)


CCC Distribution per Metric for 08D
Metric
Right Knee Angle
Left Knee Angle
Right Hip Angle
Left Hip Angle
Right Shank Sagittal Angle
Left Shank Sagittal Angle
Right Femur Sagittal Angle
Left Femur Sagittal Angle
Sagittal Trunk Angle
Frontal Trunk Angle
All Metrics
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
CCC

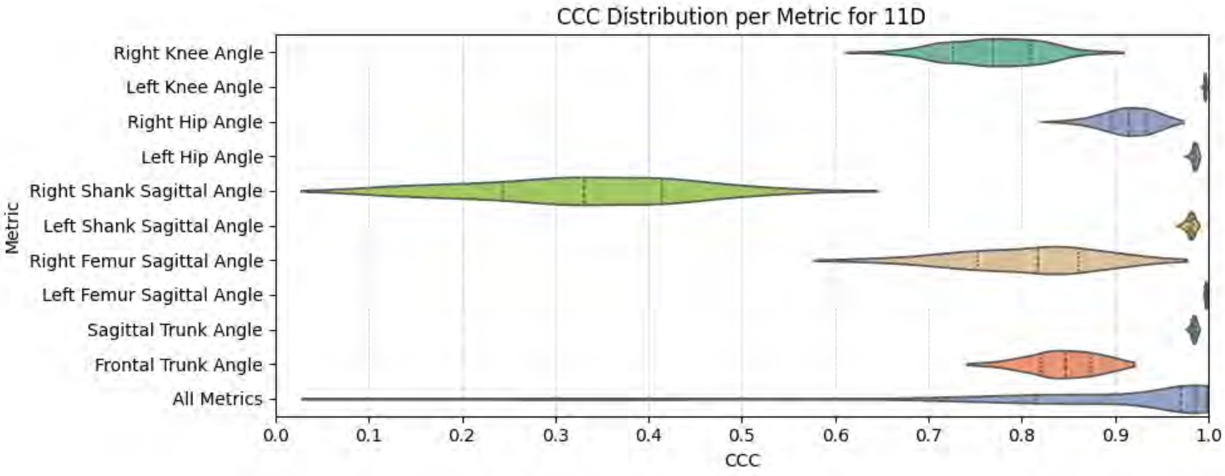
s)


CCC Distribution per Metric for 09D
Metric
Right Knee Angle
Left Knee Angle
Right Hip Angle
Left Hip Angle
Right Shank Sagittal Angle
Left Shank Sagittal Angle
Right Femur Sagittal Angle
Left Femur Sagittal Angle
Sagittal Trunk Angle
Frontal Trunk Angle
All Metrics
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
CCC

t)


CCC Distribution per Metric for 10D
Metric
Right Knee Angle
Left Knee Angle
Right Hip Angle
Left Hip Angle
Right Shank Sagittal Angle
Left Shank Sagittal Angle
Right Femur Sagittal Angle
Left Femur Sagittal Angle
Sagittal Trunk Angle
Frontal Trunk Angle
All Metrics
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
CCC

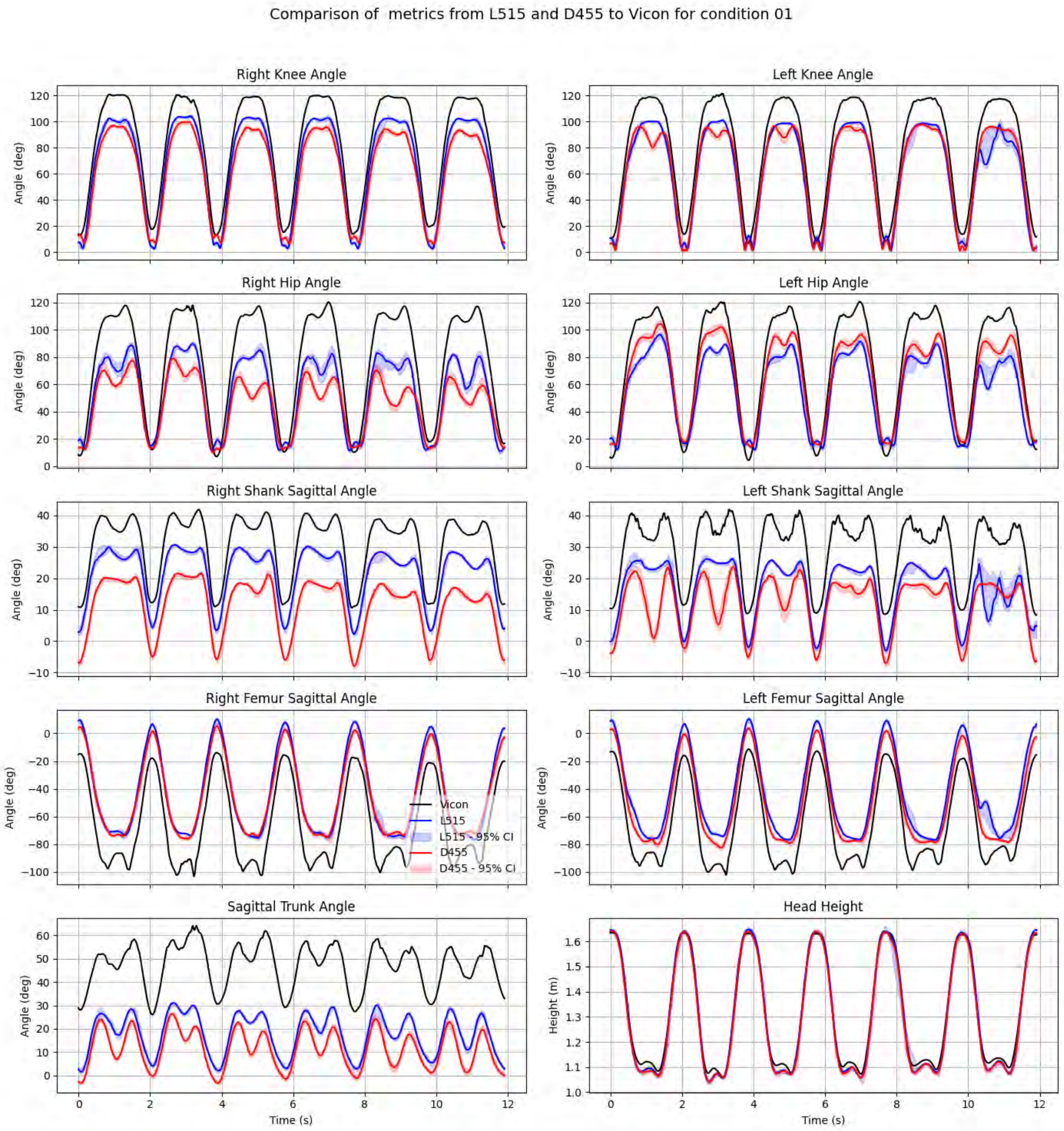
u)



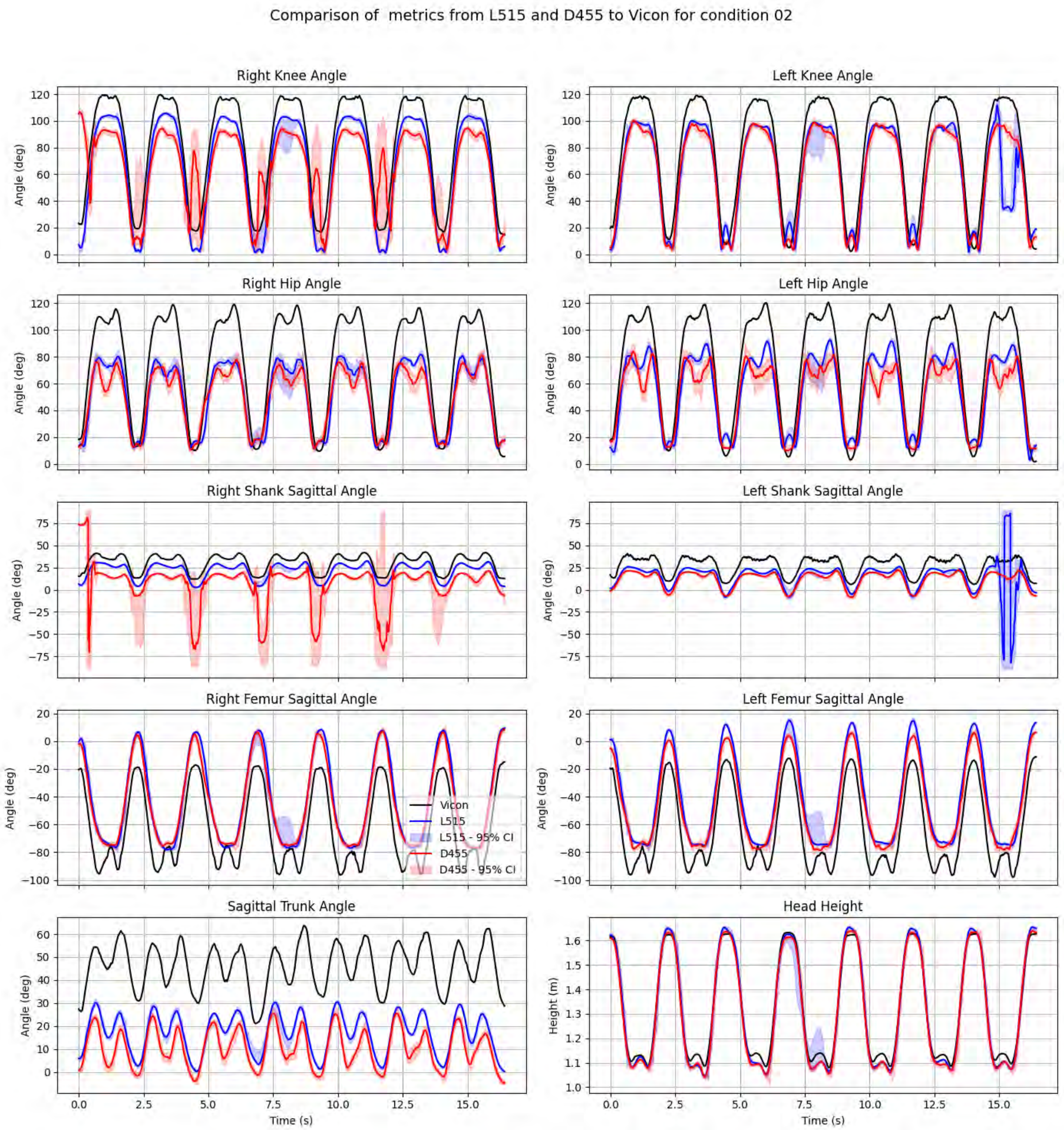
v)

Figure A.1 Violin plot CCC for all metrics per recording. Recordings from the L515 are from a) to k) and ones from the D455 are from l) to v)

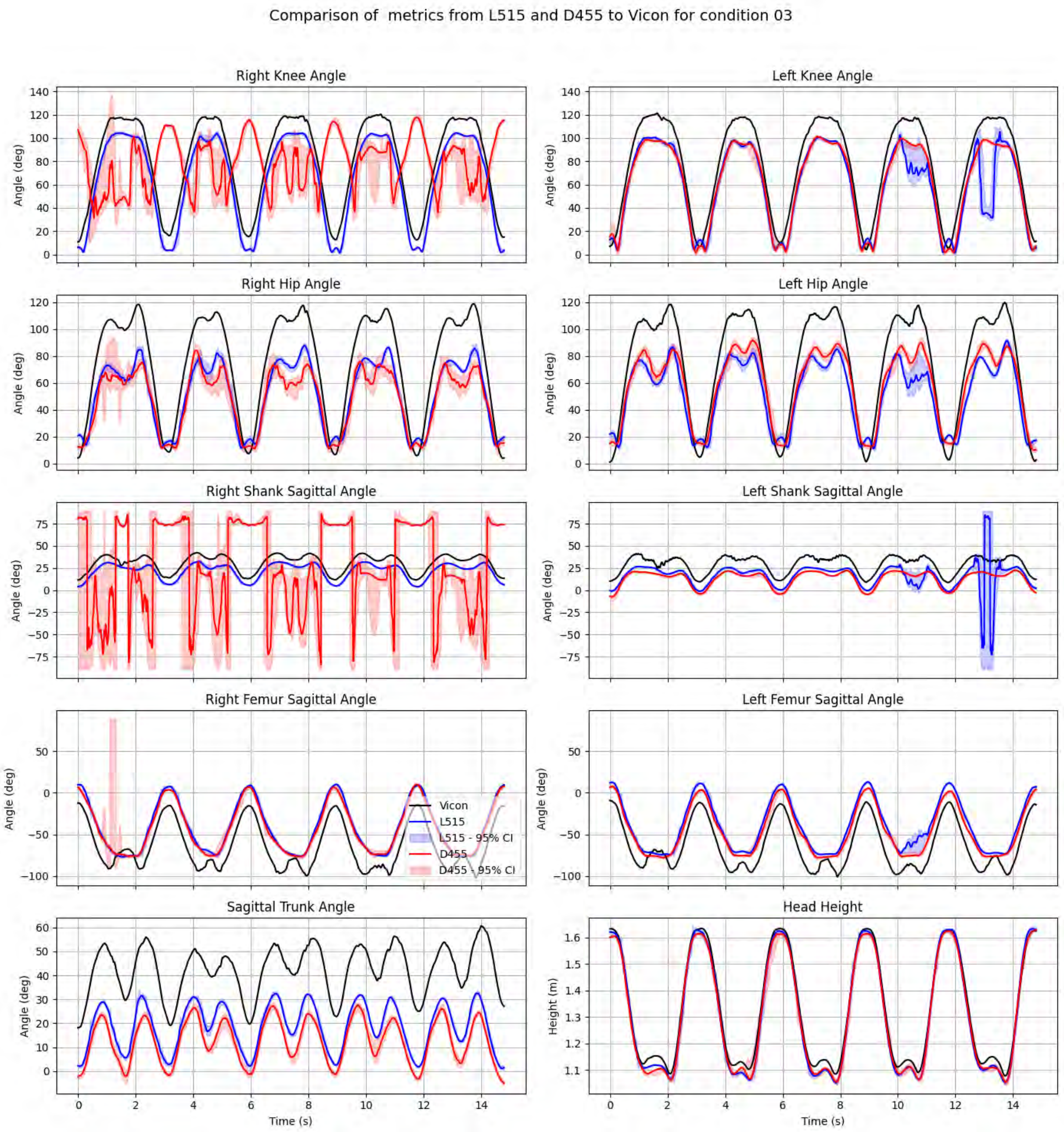
APPENDIX B COMPLETE ACCURACY RESULTS



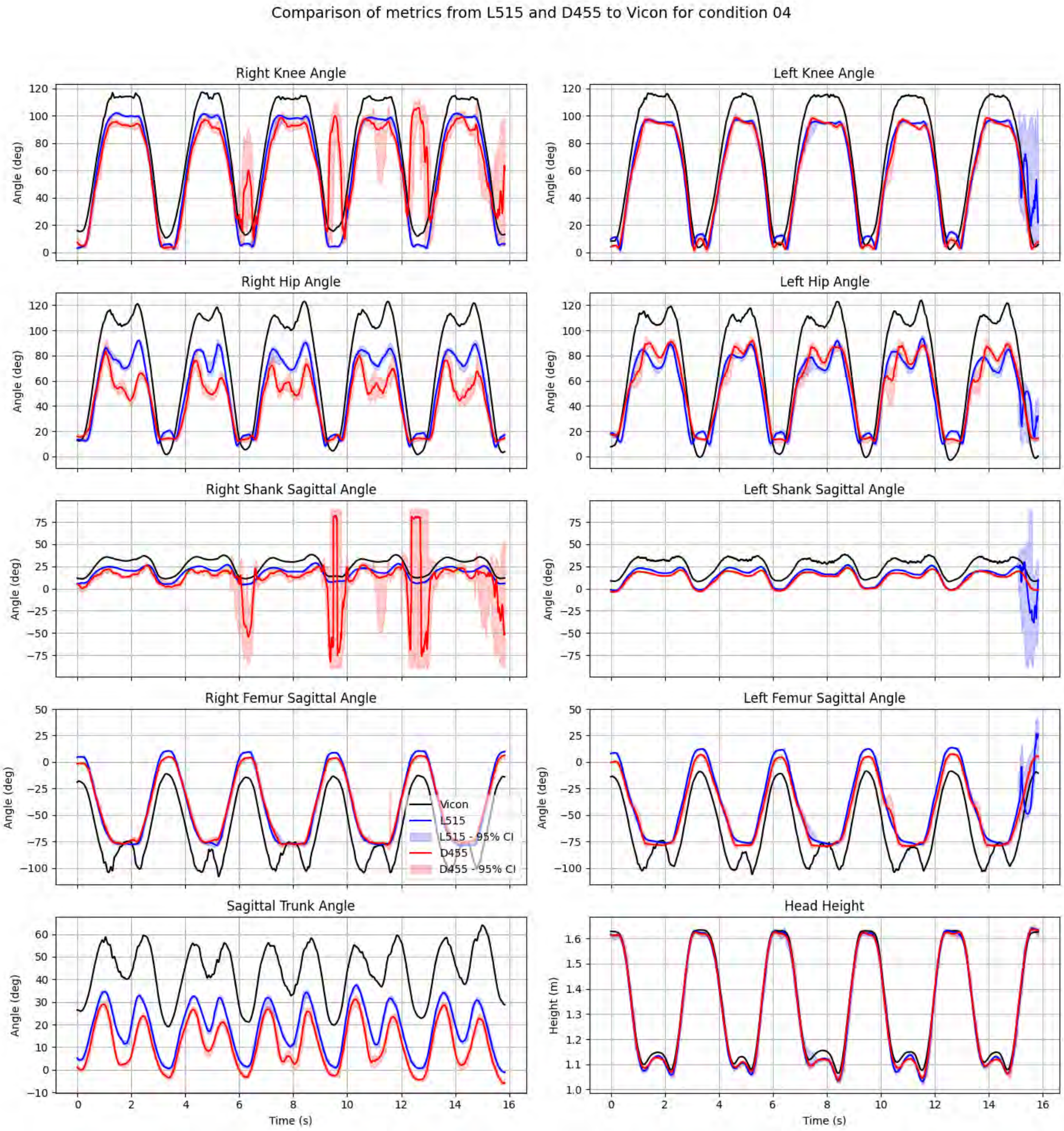
a)


Comparison of metrics from L515 and D455 to Vicon for condition 02
Right Knee Angle
Left Knee Angle
Right Hip Angle
Left Hip Angle
Right Shank Sagittal Angle
Left Shank Sagittal Angle
Right Femur Sagittal Angle
Left Femur Sagittal Angle
Sagittal Trunk Angle
Head Height
Angle (deg)
Height (m)
Time (s)
0.0
2.5
5.0
7.5
10.0
12.5
15.0
Vicon
L515
L515 - 95% CI
D455
D455 - 95% CI

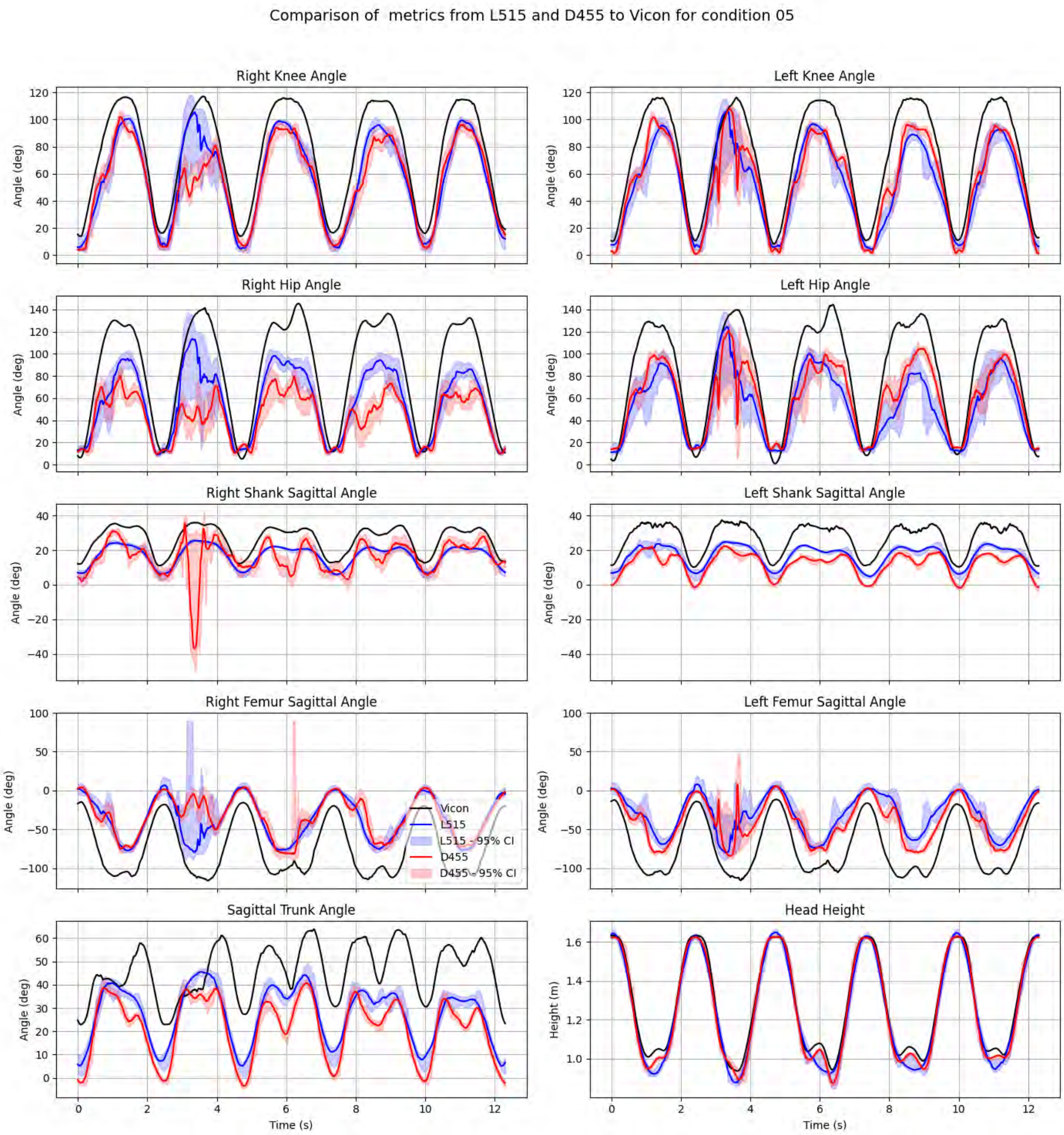
b)


Comparison of metrics from L515 and D455 to Vicon for condition 03
Right Knee Angle
Left Knee Angle
Right Hip Angle
Left Hip Angle
Right Shank Sagittal Angle
Left Shank Sagittal Angle
Right Femur Sagittal Angle
Left Femur Sagittal Angle
Sagittal Trunk Angle
Head Height
Angle (deg)
Height (m)
Time (s)
0
2
4
6
8
10
12
14
Vicon
L515
L515 - 95% CI
D455
D455 - 95% CI

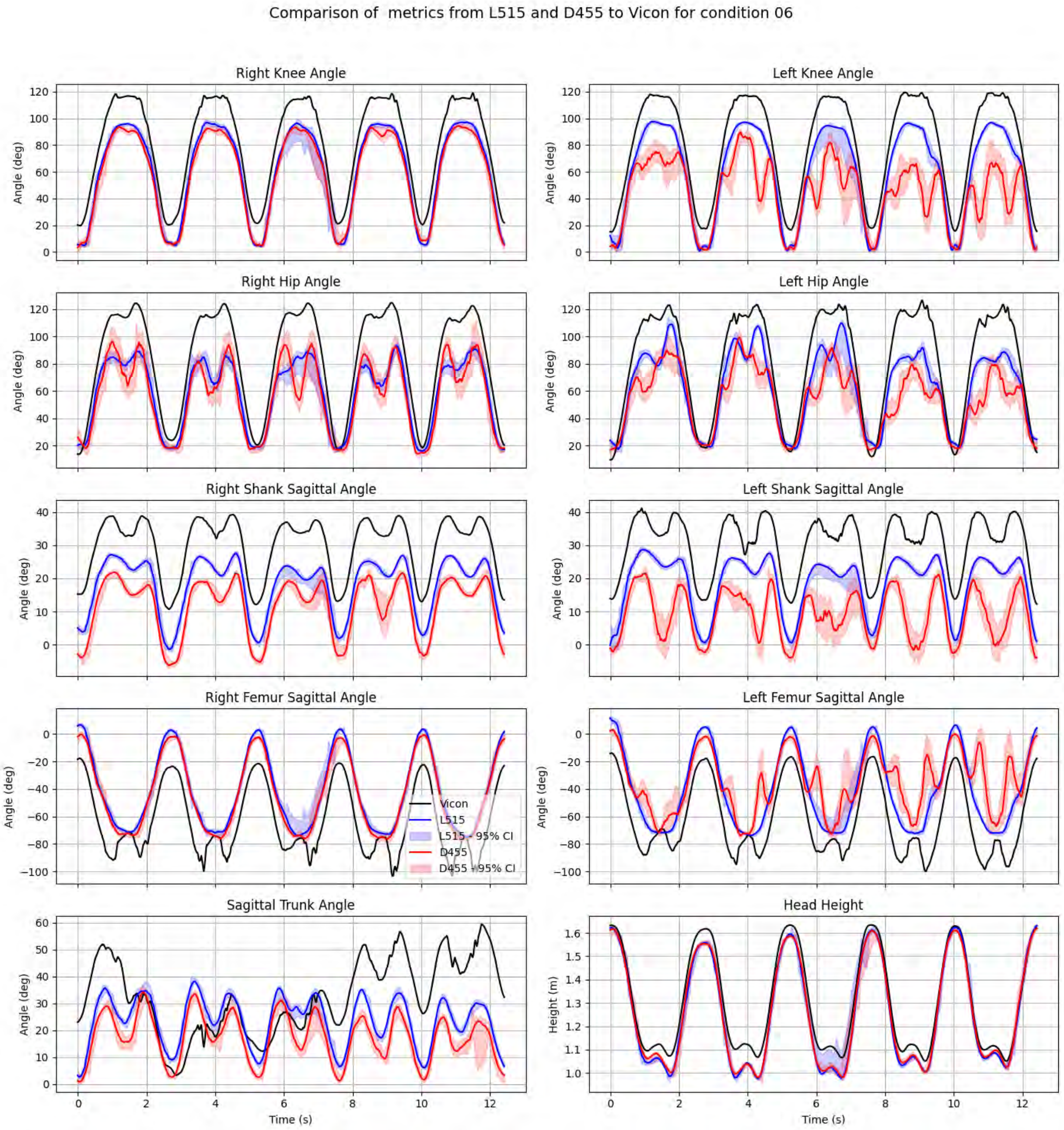
c)


Comparison of metrics from L515 and D455 to Vicon for condition 04
Right Knee Angle
Left Knee Angle
Right Hip Angle
Left Hip Angle
Right Shank Sagittal Angle
Left Shank Sagittal Angle
Right Femur Sagittal Angle
Left Femur Sagittal Angle
Sagittal Trunk Angle
Head Height
Angle (deg)
Height (m)
Time (s)
0
2
4
6
8
10
12
14
16
Vicon
L515
L515 - 95% CI
D455
D455 - 95% CI

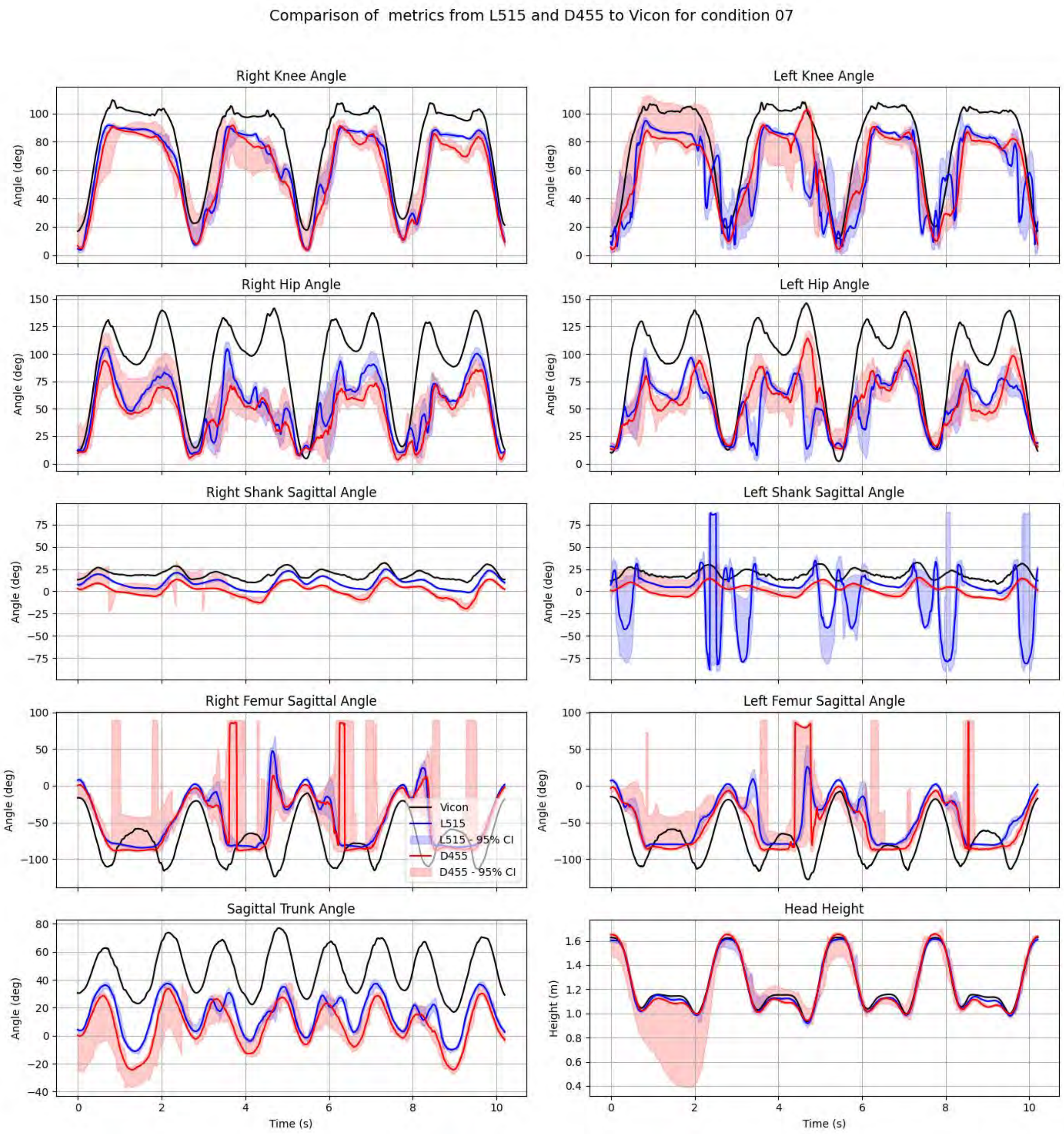
d)


Comparison of metrics from L515 and D455 to Vicon for condition 05
Right Knee Angle
Left Knee Angle
Right Hip Angle
Left Hip Angle
Right Shank Sagittal Angle
Left Shank Sagittal Angle
Right Femur Sagittal Angle
Left Femur Sagittal Angle
Sagittal Trunk Angle
Head Height
Angle (deg)
Height (m)
Time (s)
0
2
4
6
8
10
12
120
100
80
60
40
20
140
−20
−40
50
−50
−100
1.6
1.4
1.2
1.0
Vicon
L515
L515 - 95% CI
D455
D455 - 95% CI

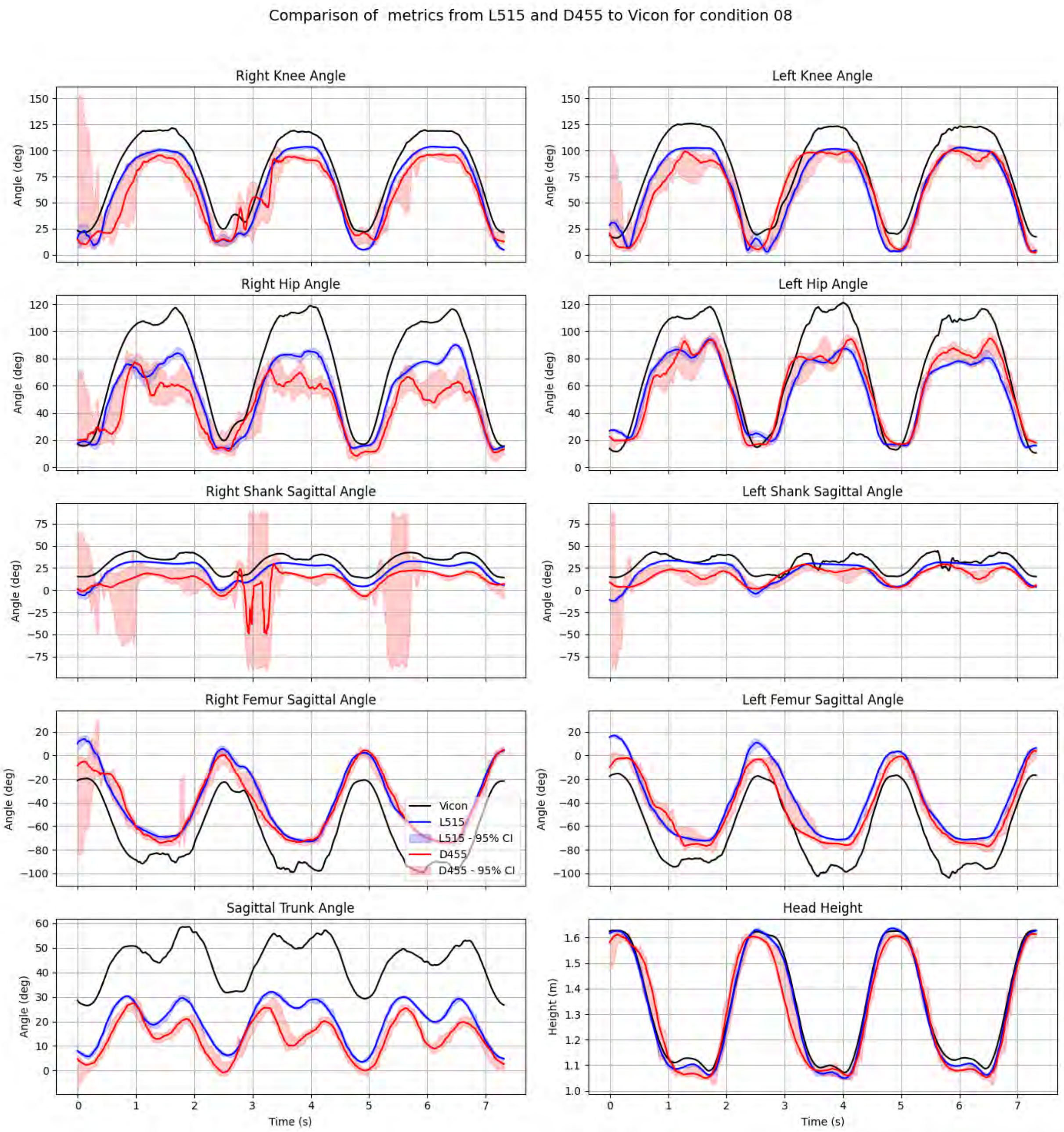
e)


Comparison of metrics from L515 and D455 to Vicon for condition 06
Right Knee Angle
Left Knee Angle
Right Hip Angle
Left Hip Angle
Right Shank Sagittal Angle
Left Shank Sagittal Angle
Right Femur Sagittal Angle
Left Femur Sagittal Angle
Sagittal Trunk Angle
Head Height
Angle (deg)
Height (m)
Time (s)
Vicon
L515
L515 - 95% CI
D455
D455 - 95% CI

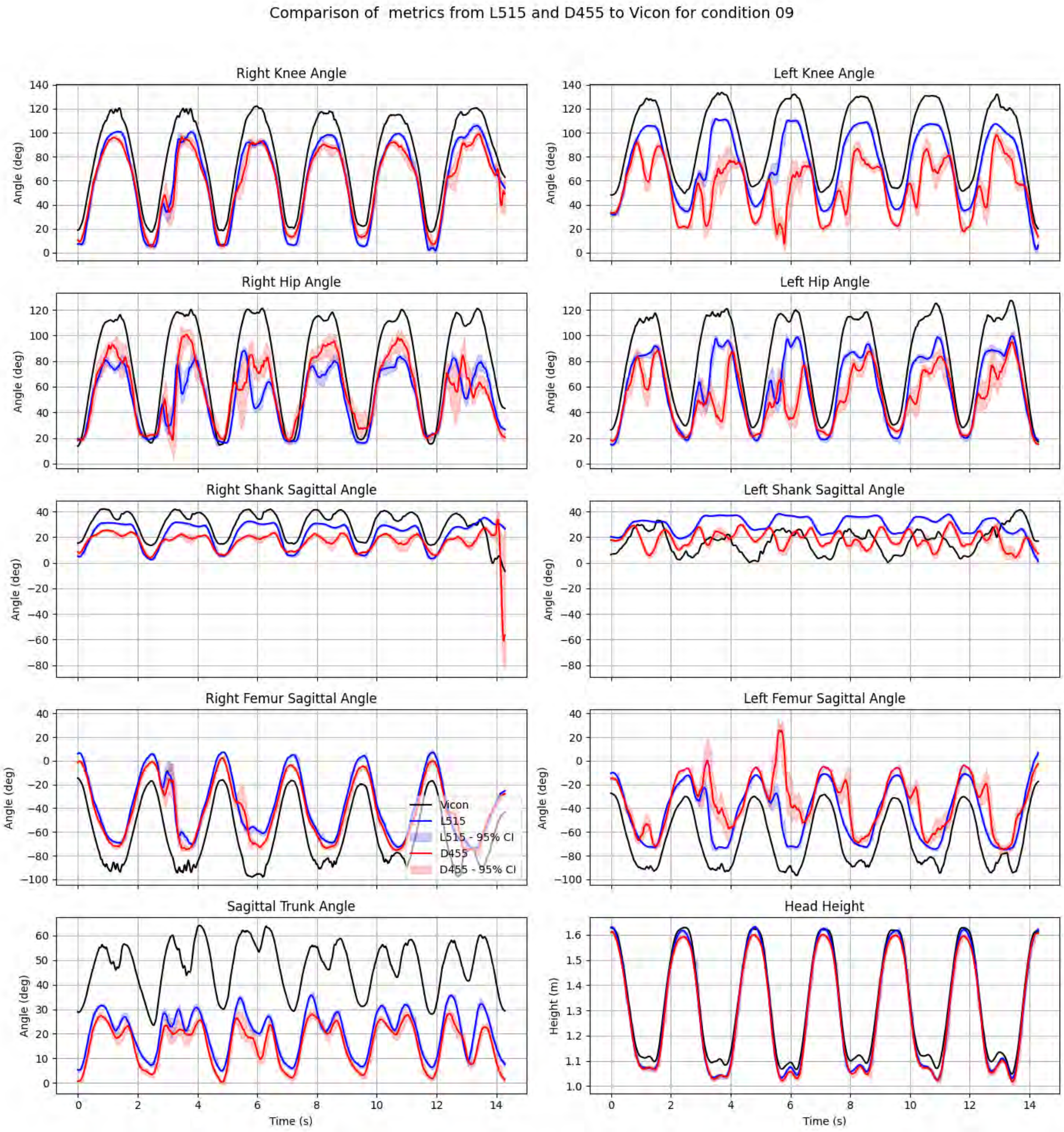
f)


Comparison of metrics from L515 and D455 to Vicon for condition 07
Right Knee Angle
Left Knee Angle
Right Hip Angle
Left Hip Angle
Right Shank Sagittal Angle
Left Shank Sagittal Angle
Right Femur Sagittal Angle
Left Femur Sagittal Angle
Sagittal Trunk Angle
Head Height
Angle (deg)
Height (m)
Time (s)
Vicon
L515
L515 - 95% CI
D455
D455 - 95% CI

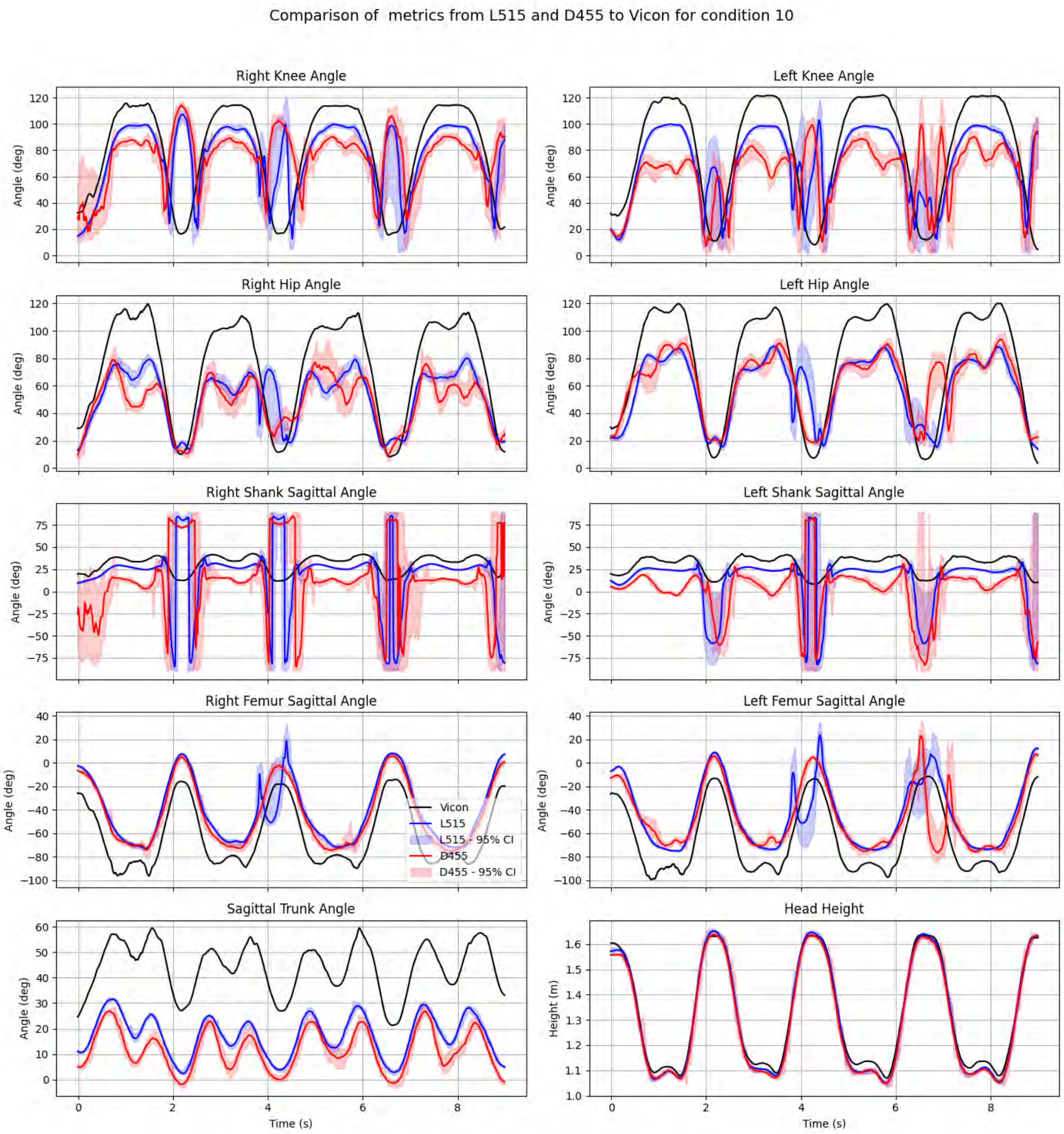
g)


Comparison of metrics from L515 and D455 to Vicon for condition 08
Right Knee Angle
Left Knee Angle
Right Hip Angle
Left Hip Angle
Right Shank Sagittal Angle
Left Shank Sagittal Angle
Right Femur Sagittal Angle
Left Femur Sagittal Angle
Sagittal Trunk Angle
Head Height
Angle (deg)
Height (m)
Time (s)
Vicon
L515
L515 - 95% CI
D455
D455 - 95% CI

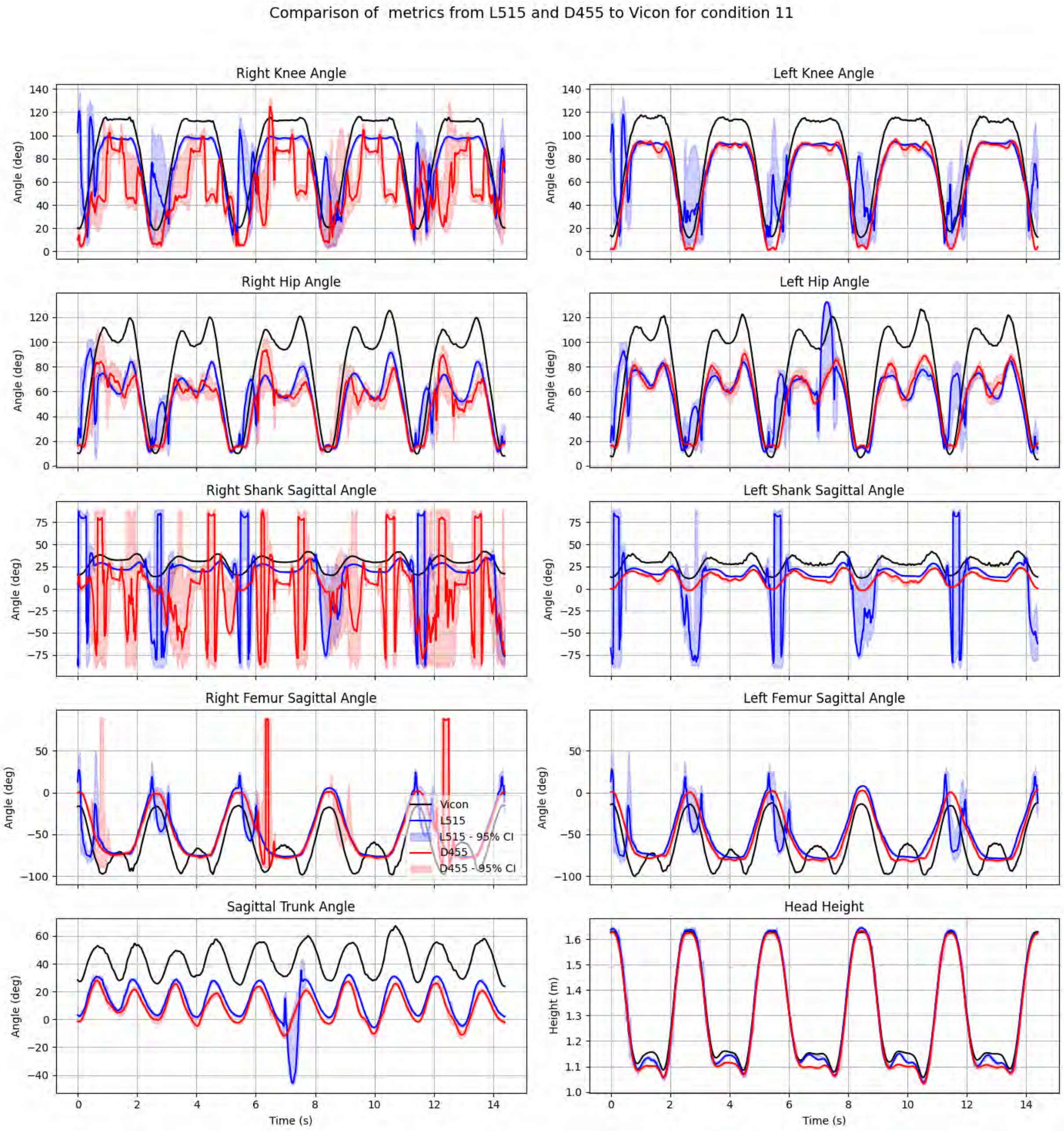
h)


Comparison of metrics from L515 and D455 to Vicon for condition 09
Right Knee Angle
Left Knee Angle
Right Hip Angle
Left Hip Angle
Right Shank Sagittal Angle
Left Shank Sagittal Angle
Right Femur Sagittal Angle
Left Femur Sagittal Angle
Sagittal Trunk Angle
Head Height
Angle (deg)
Height (m)
Time (s)
0
2
4
6
8
10
12
14
Vicon
L515
L515 - 95% CI
D455
D455 - 95% CI

i)

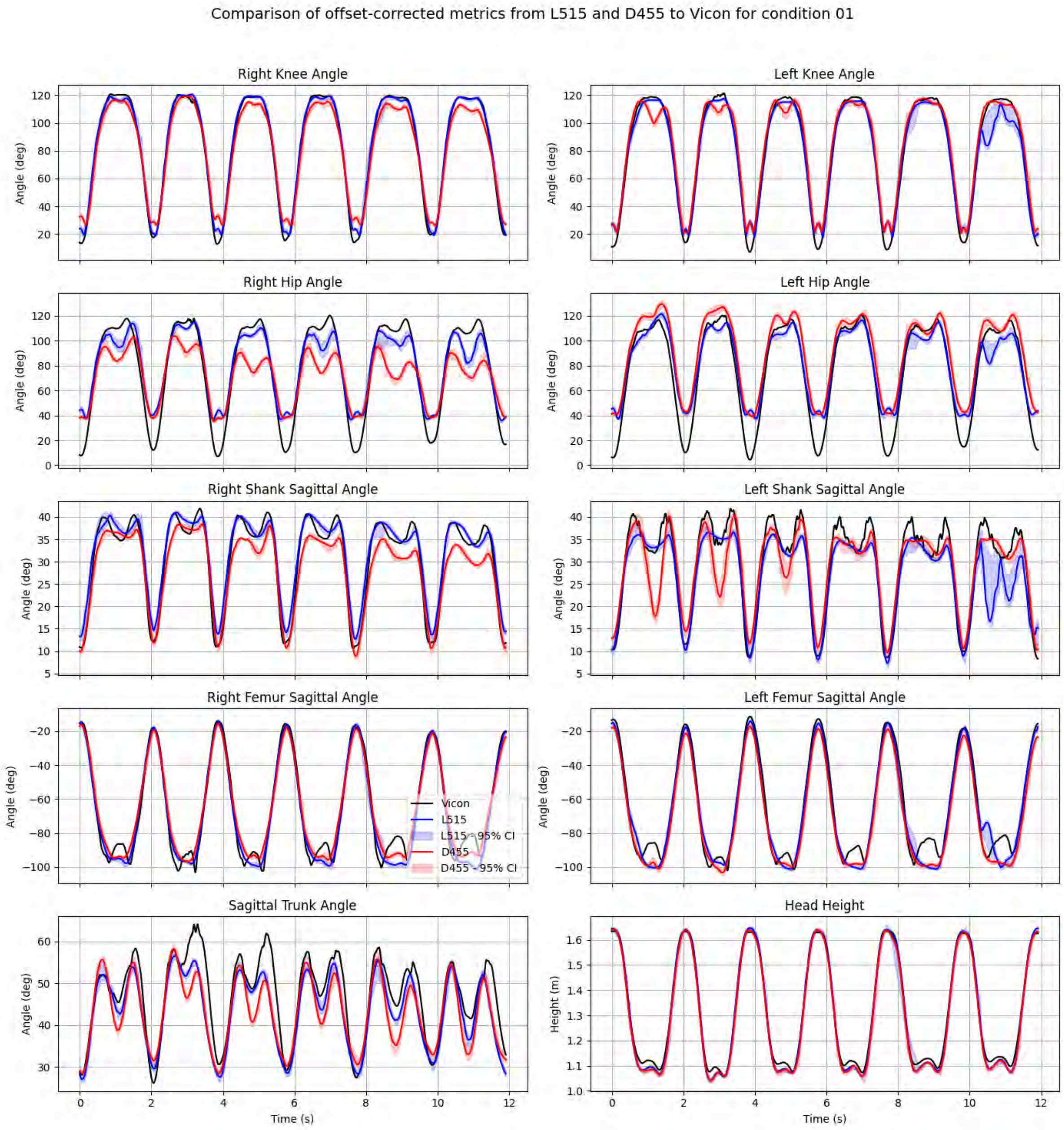

Comparison of metrics from L515 and D455 to Vicon for condition 10
Right Knee Angle
Left Knee Angle
Right Hip Angle
Left Hip Angle
Right Shank Sagittal Angle
Left Shank Sagittal Angle
Right Femur Sagittal Angle
Left Femur Sagittal Angle
Sagittal Trunk Angle
Head Height
Angle (deg)
Height (m)
Time (s)
Vicon
L515
L515 - 95% CI
D455
D455 - 95% CI
0
2
4
6
8

j)

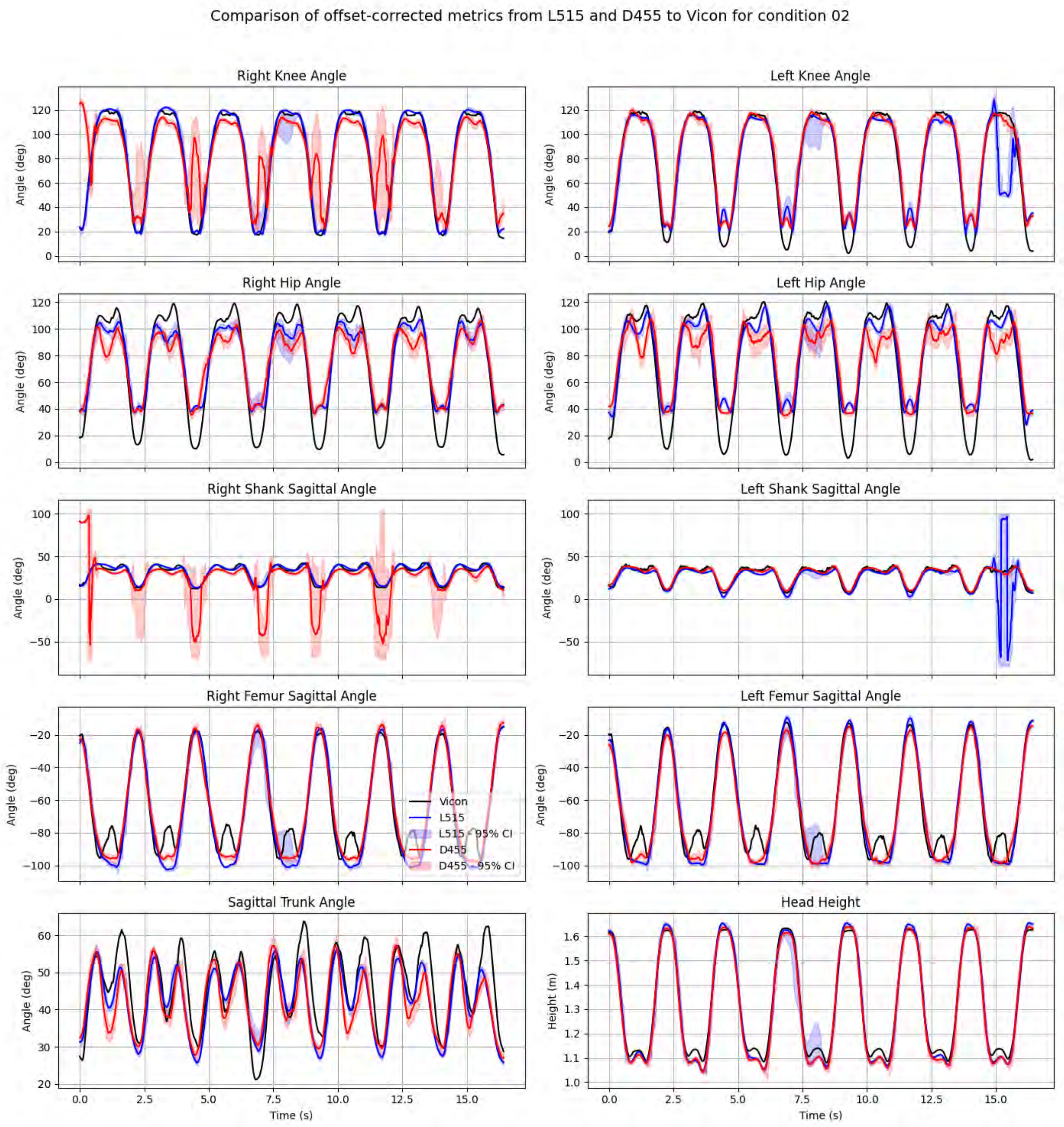


k)

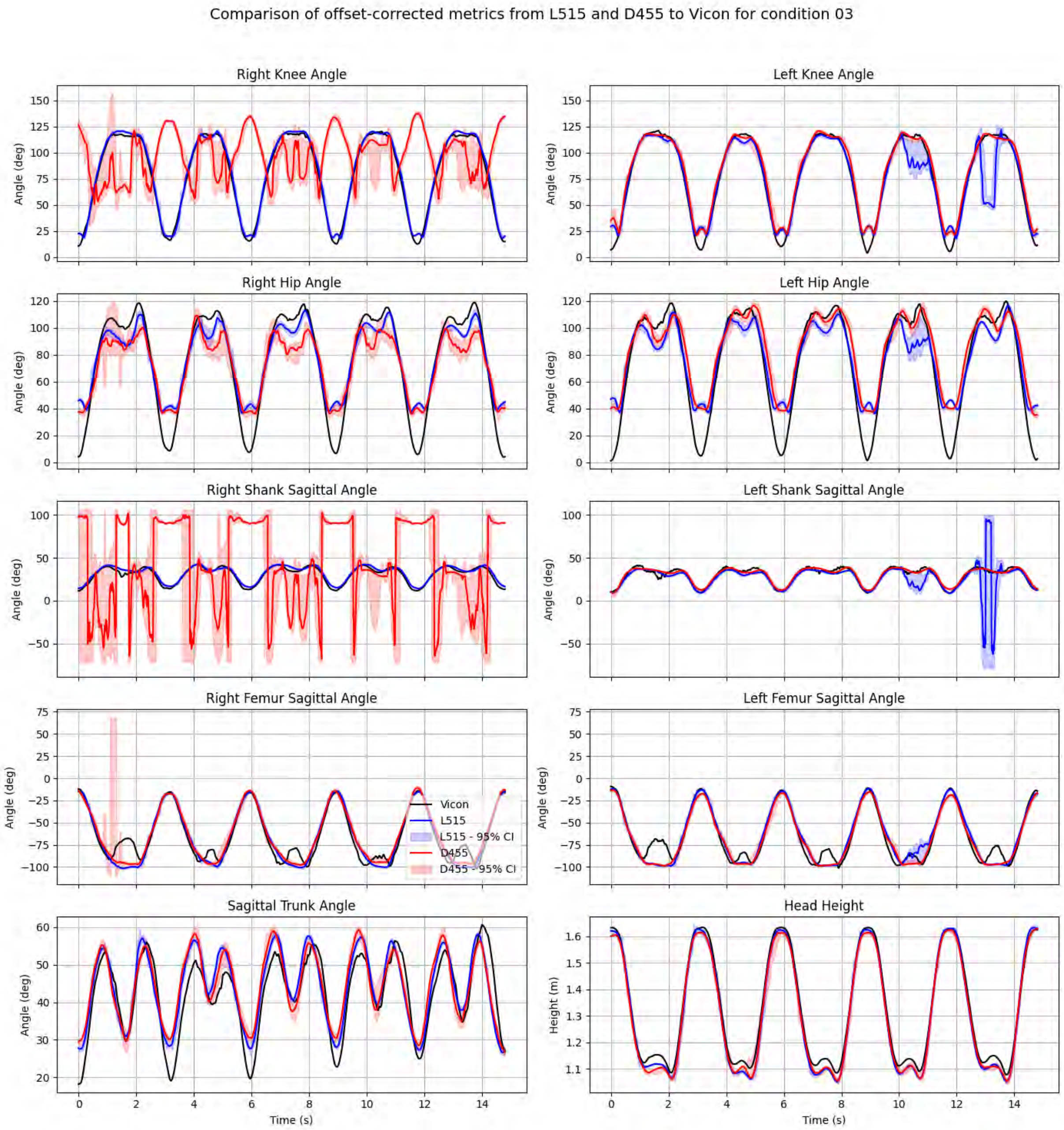
Figure B.1 Comparison of metrics for the 11 conditions between estimated data and the ground truth. Conditions 01 to 11 are in order from a) to k)


Comparison of offset-corrected metrics from L515 and D455 to Vicon for condition 01
Right Knee Angle
Left Knee Angle
Right Hip Angle
Left Hip Angle
Right Shank Sagittal Angle
Left Shank Sagittal Angle
Right Femur Sagittal Angle
Left Femur Sagittal Angle
Sagittal Trunk Angle
Head Height
Angle (deg)
Height (m)
Time (s)
0
2
4
6
8
10
12
Vicon
L515
L515 - 95% CI
D455
D455 - 95% CI

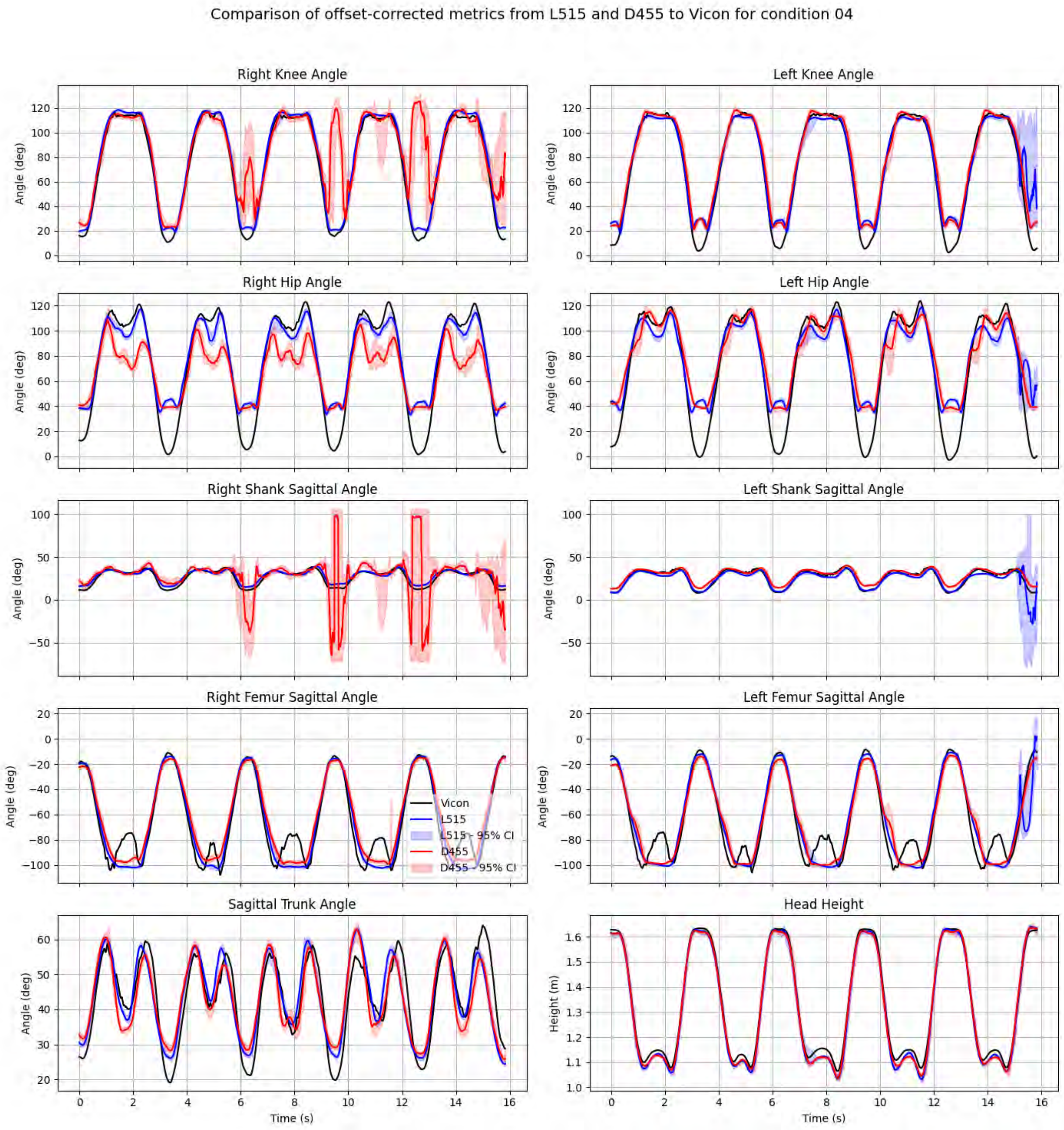
a)


Comparison of offset-corrected metrics from L515 and D455 to Vicon for condition 02
Right Knee Angle
Left Knee Angle
Right Hip Angle
Left Hip Angle
Right Shank Sagittal Angle
Left Shank Sagittal Angle
Right Femur Sagittal Angle
Left Femur Sagittal Angle
Sagittal Trunk Angle
Head Height
Angle (deg)
Height (m)
Time (s)
0.0
2.5
5.0
7.5
10.0
12.5
15.0
Vicon
L515
L515 - 95% CI
D455
D455 - 95% CI

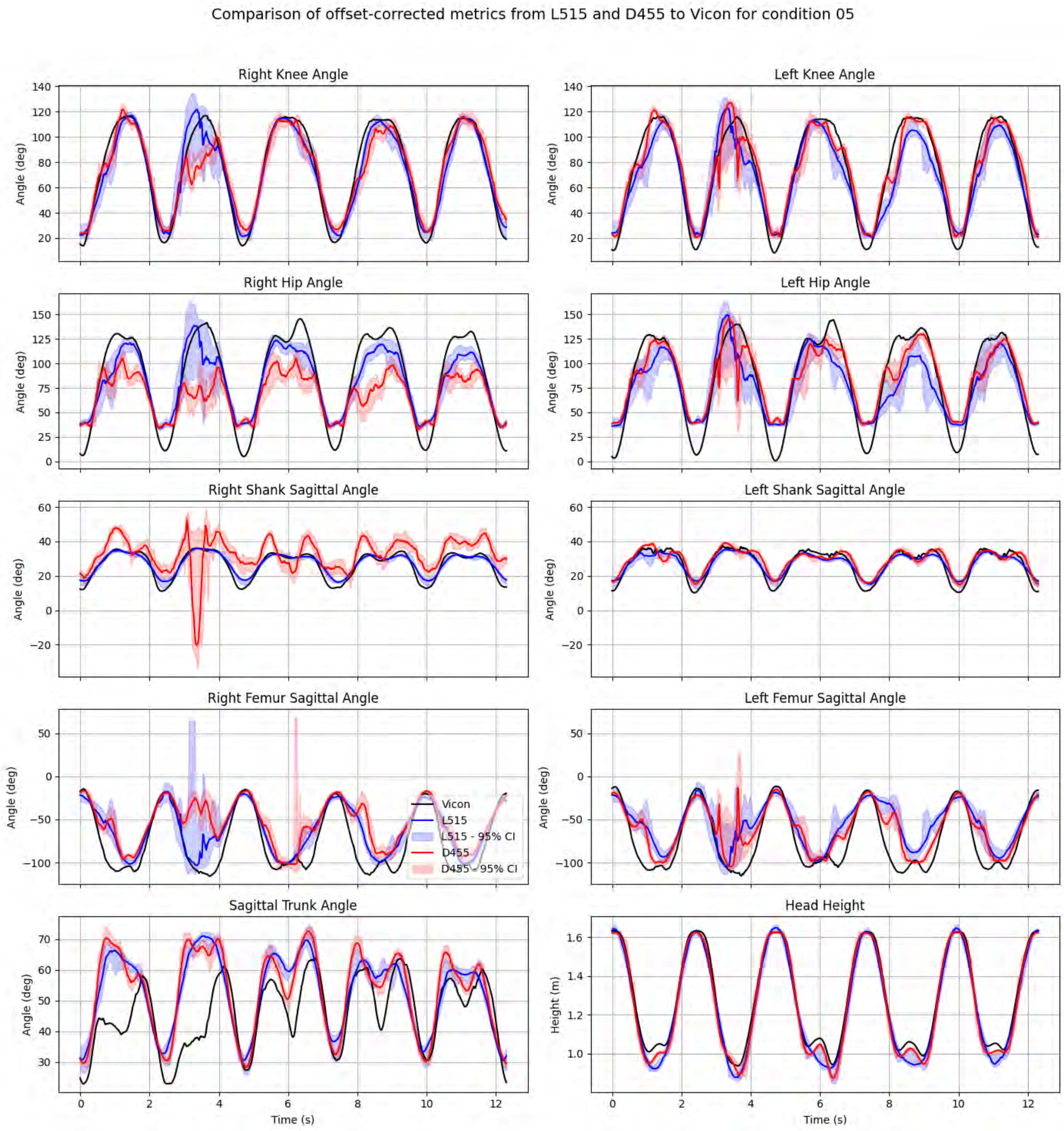
b)


Comparison of offset-corrected metrics from L515 and D455 to Vicon for condition 03
Right Knee Angle
Left Knee Angle
Right Hip Angle
Left Hip Angle
Right Shank Sagittal Angle
Left Shank Sagittal Angle
Right Femur Sagittal Angle
Left Femur Sagittal Angle
Sagittal Trunk Angle
Head Height
Angle (deg)
Height (m)
Time (s)
0
2
4
6
8
10
12
14
Vicon
L515
L515 - 95% CI
D455
D455 - 95% CI

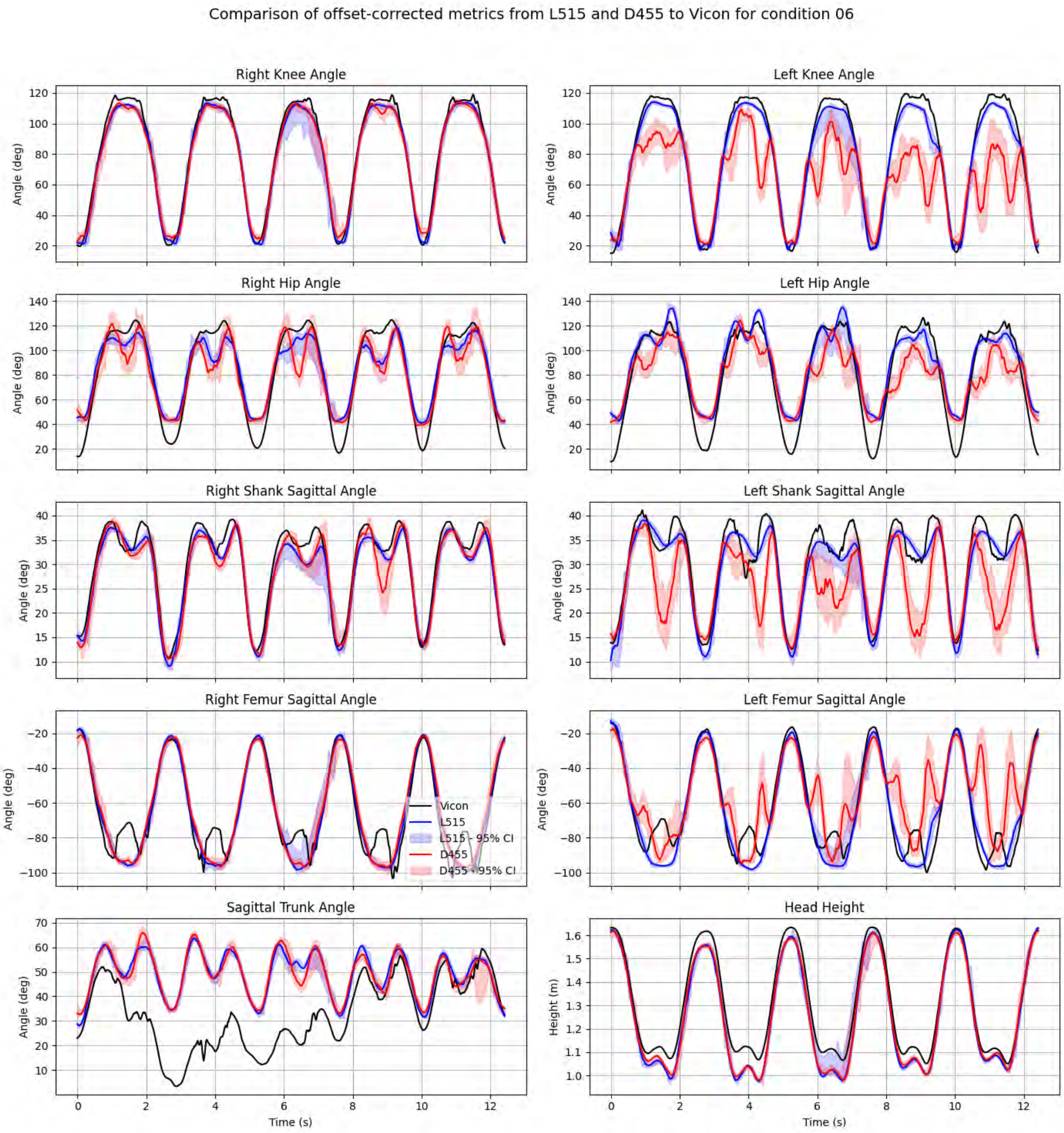
c)


Comparison of offset-corrected metrics from L515 and D455 to Vicon for condition 04
Right Knee Angle
Left Knee Angle
Right Hip Angle
Left Hip Angle
Right Shank Sagittal Angle
Left Shank Sagittal Angle
Right Femur Sagittal Angle
Left Femur Sagittal Angle
Sagittal Trunk Angle
Head Height
Angle (deg)
Height (m)
Time (s)
0
2
4
6
8
10
12
14
16
Vicon
L515
L515 - 95% CI
D455
D455 - 95% CI

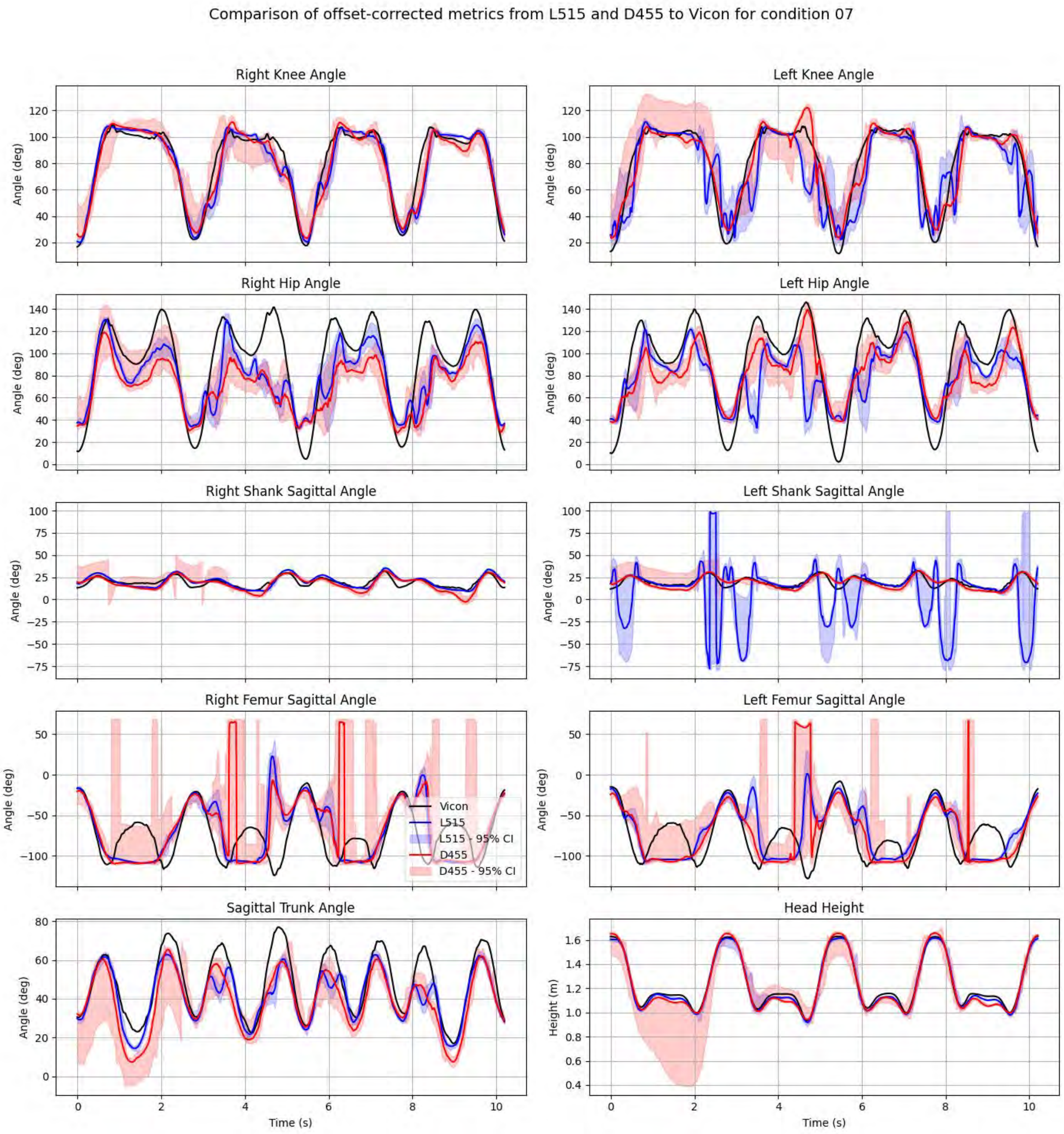
d)


Comparison of offset-corrected metrics from L515 and D455 to Vicon for condition 05
Right Knee Angle
Left Knee Angle
Right Hip Angle
Left Hip Angle
Right Shank Sagittal Angle
Left Shank Sagittal Angle
Right Femur Sagittal Angle
Left Femur Sagittal Angle
Sagittal Trunk Angle
Head Height
Angle (deg)
Height (m)
Time (s)
Vicon
L515
L515 - 95% CI
D455
D455 - 95% CI

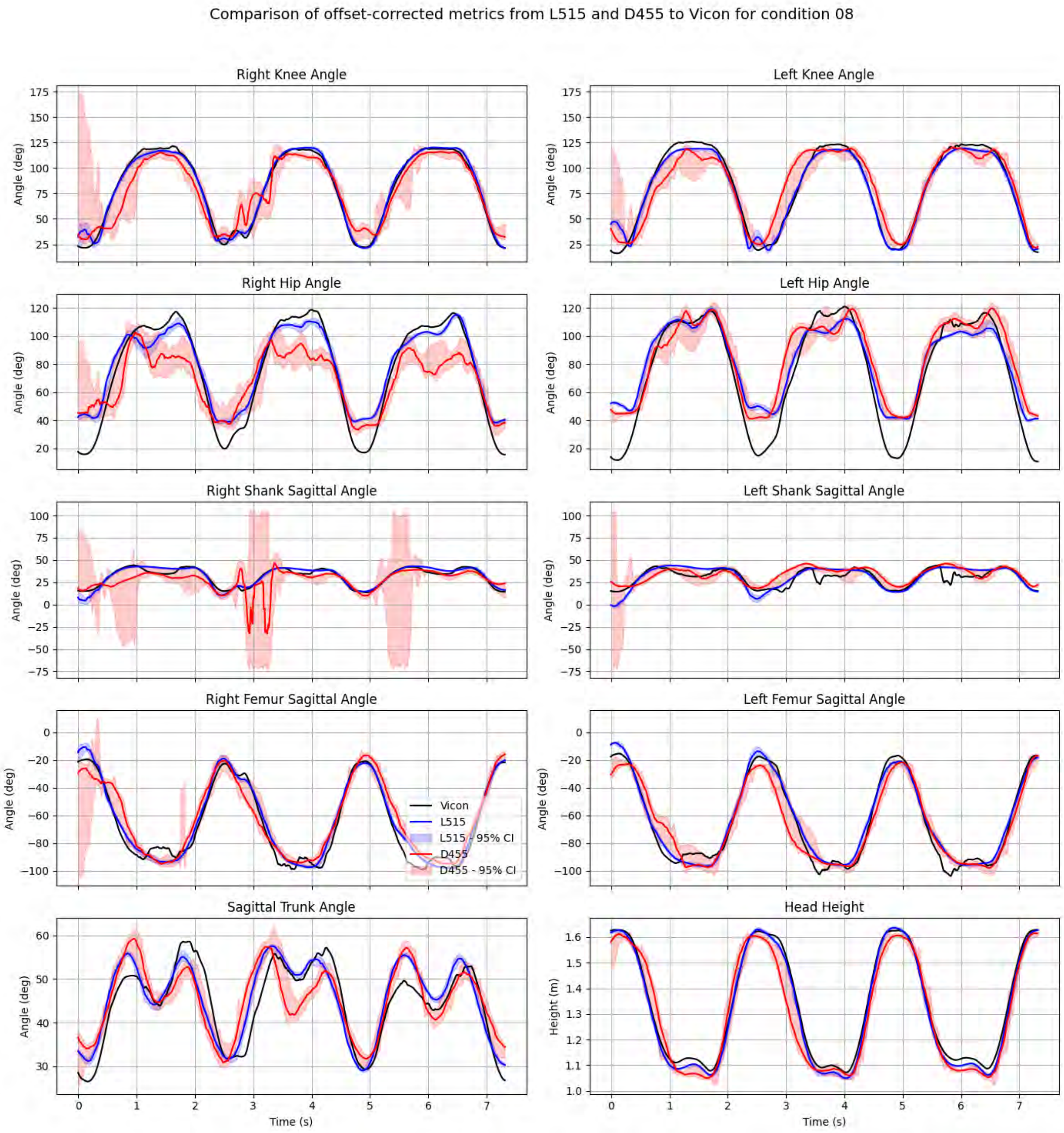
e)


Comparison of offset-corrected metrics from L515 and D455 to Vicon for condition 06
Right Knee Angle
Left Knee Angle
Right Hip Angle
Left Hip Angle
Right Shank Sagittal Angle
Left Shank Sagittal Angle
Right Femur Sagittal Angle
Left Femur Sagittal Angle
Sagittal Trunk Angle
Head Height
Angle (deg)
Height (m)
Time (s)
0
2
4
6
8
10
12
Vicon
L515
L515 - 95% CI
D455
D455 - 95% CI

f)


Comparison of offset-corrected metrics from L515 and D455 to Vicon for condition 07
Right Knee Angle
Left Knee Angle
Right Hip Angle
Left Hip Angle
Right Shank Sagittal Angle
Left Shank Sagittal Angle
Right Femur Sagittal Angle
Left Femur Sagittal Angle
Sagittal Trunk Angle
Head Height
Angle (deg)
Height (m)
Time (s)
Vicon
L515
L515 - 95% CI
D455
D455 - 95% CI

g)


Comparison of offset-corrected metrics from L515 and D455 to Vicon for condition 08

Right Knee Angle

Left Knee Angle

Right Hip Angle

Left Hip Angle

Right Shank Sagittal Angle

Left Shank Sagittal Angle

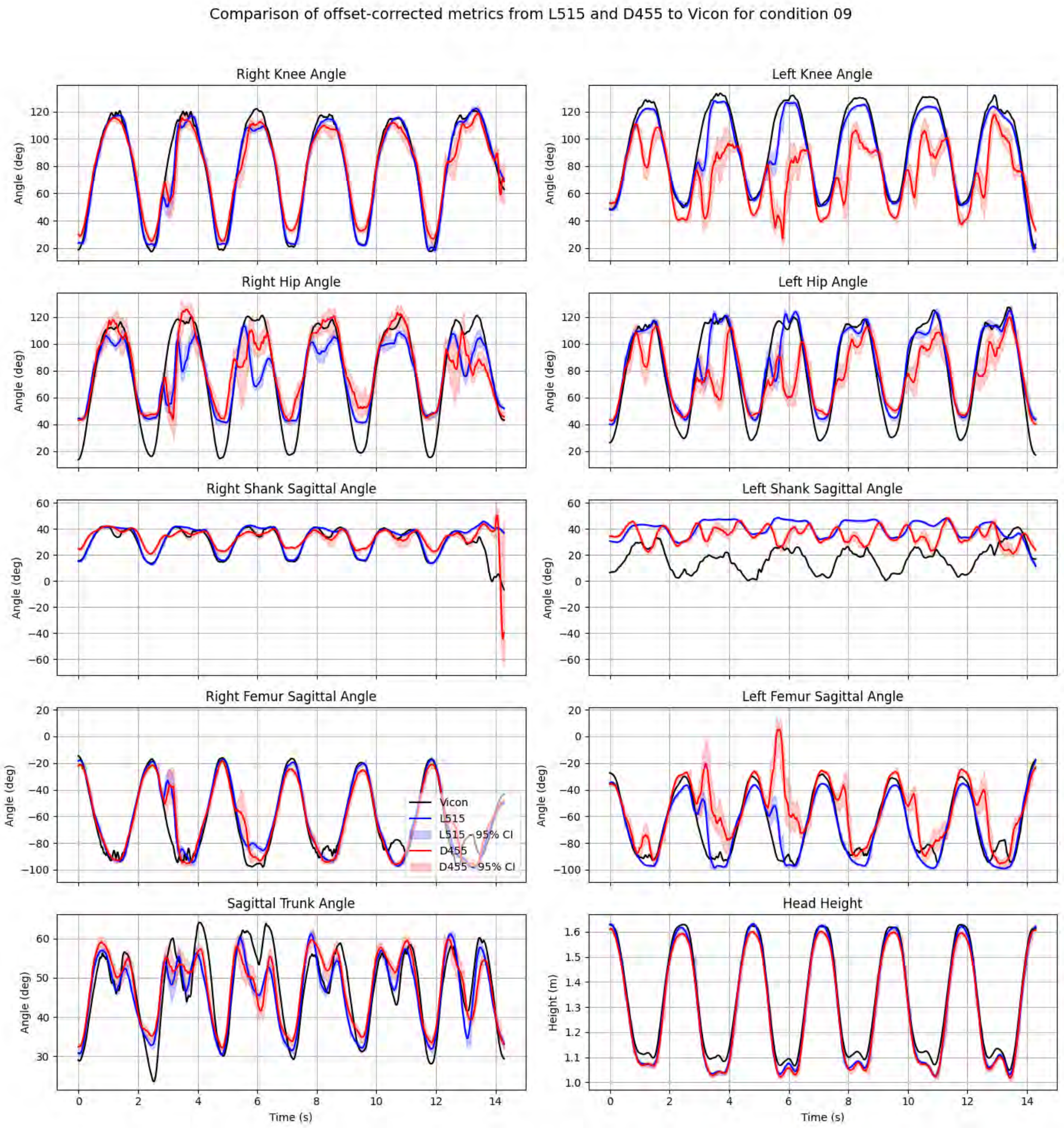
Right Femur Sagittal Angle

Left Femur Sagittal Angle

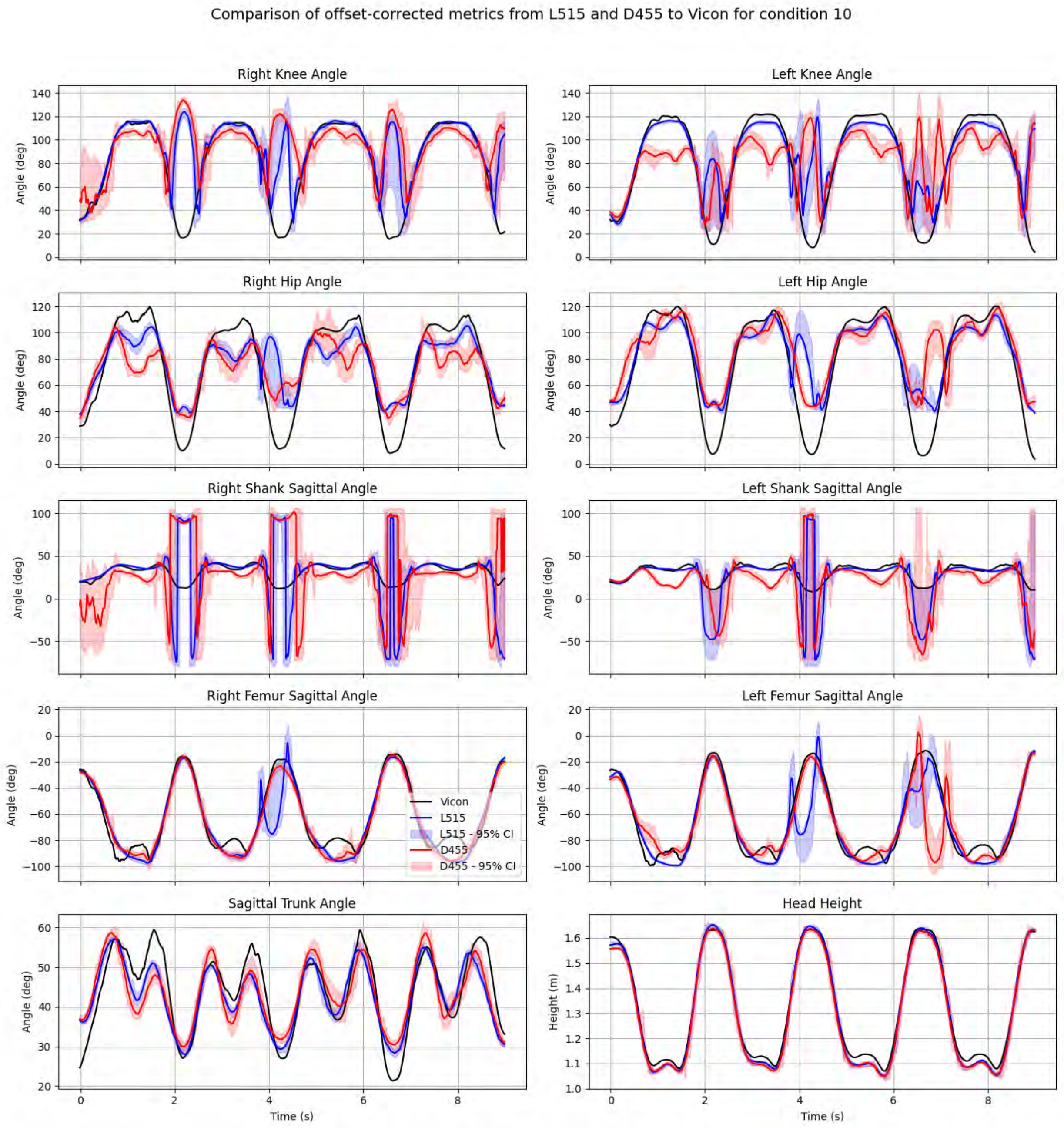
Sagittal Trunk Angle

Head Height

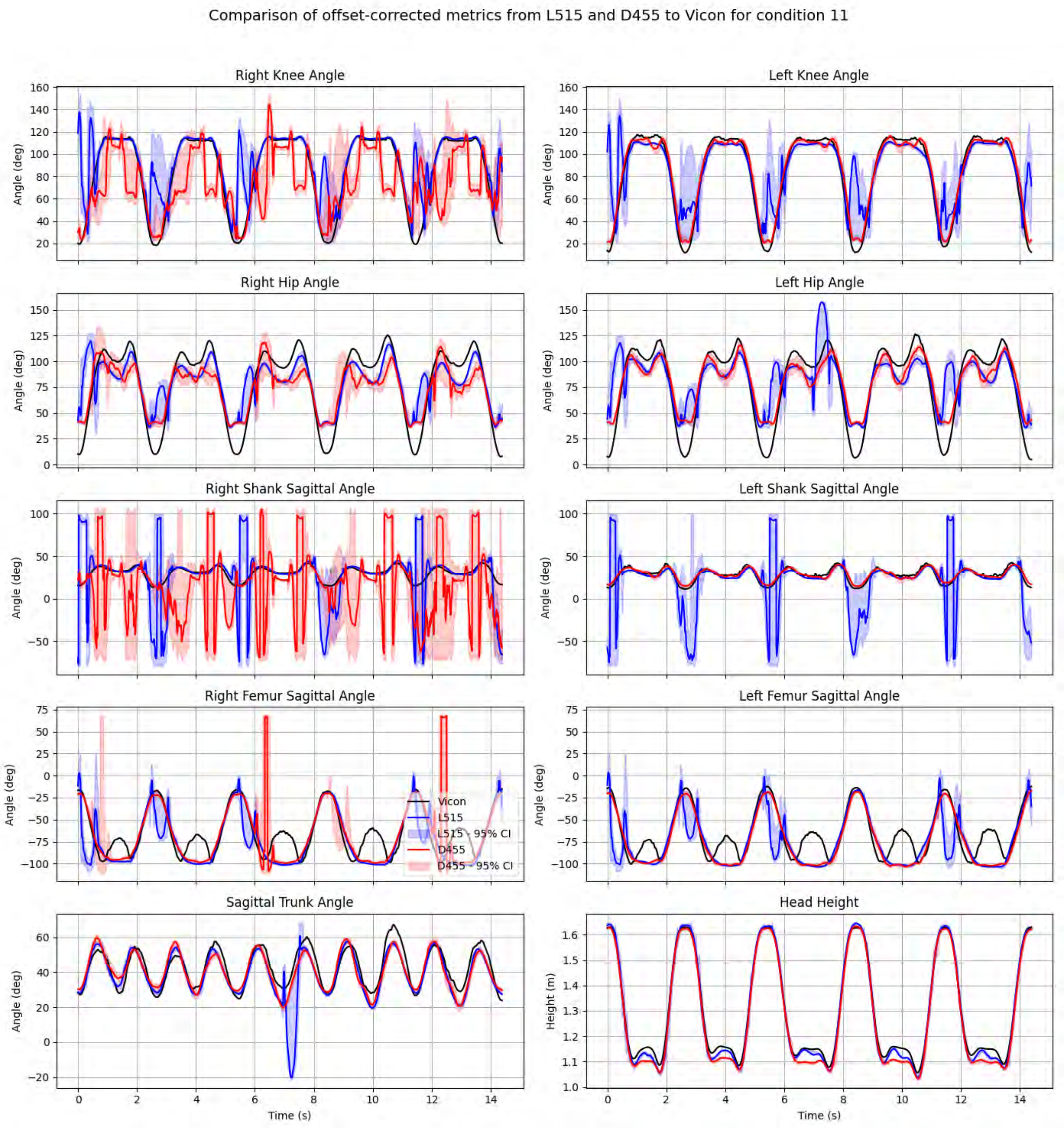
h)


Comparison of offset-corrected metrics from L515 and D455 to Vicon for condition 09
Right Knee Angle
Left Knee Angle
Right Hip Angle
Left Hip Angle
Right Shank Sagittal Angle
Left Shank Sagittal Angle
Right Femur Sagittal Angle
Left Femur Sagittal Angle
Sagittal Trunk Angle
Head Height
Angle (deg)
Height (m)
Time (s)
Vicon
L515
L515 - 95% CI
D455
D455 - 95% CI
0
2
4
6
8
10
12
14

i)

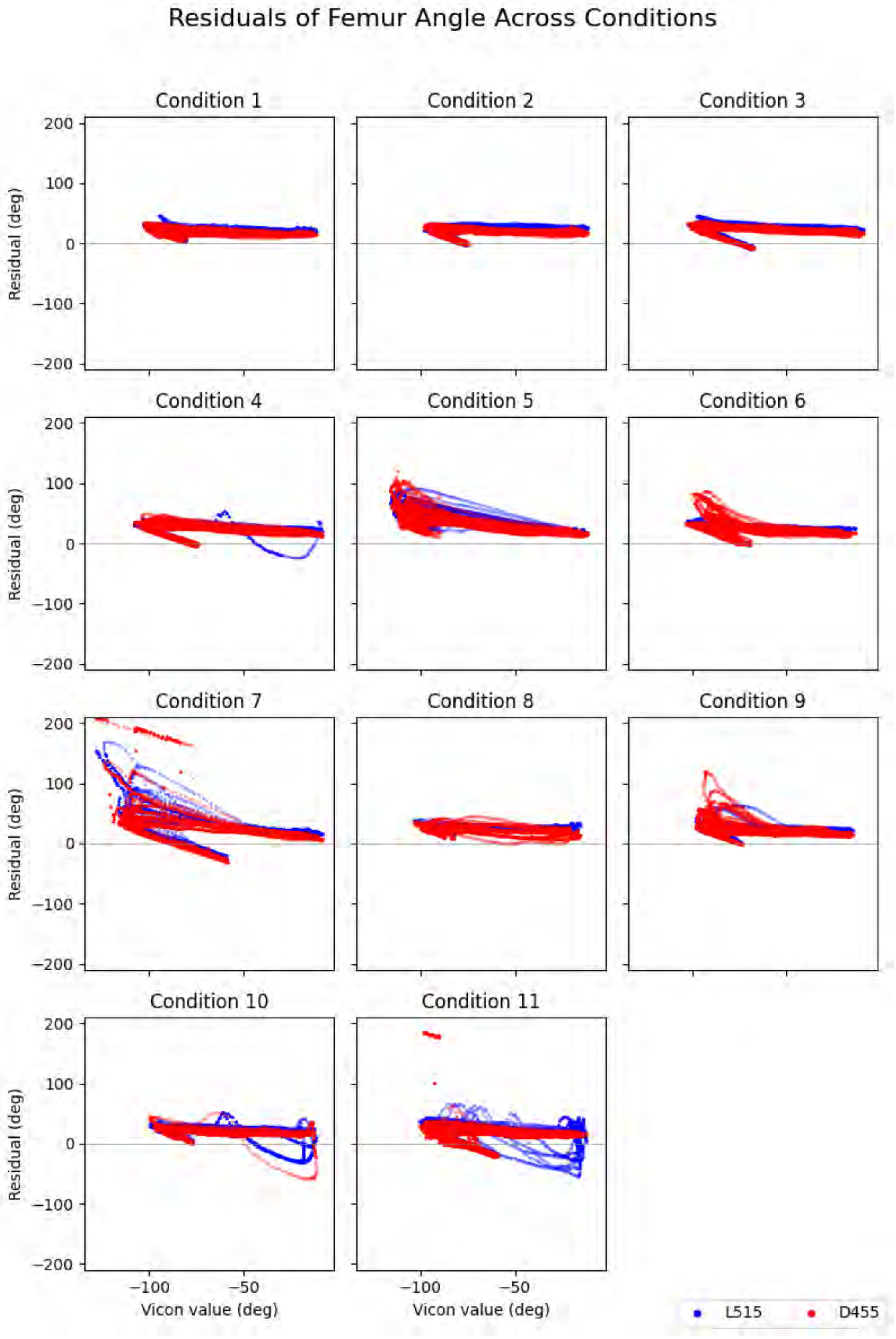

Comparison of offset-corrected metrics from L515 and D455 to Vicon for condition 10
Right Knee Angle
Left Knee Angle
Right Hip Angle
Left Hip Angle
Right Shank Sagittal Angle
Left Shank Sagittal Angle
Right Femur Sagittal Angle
Left Femur Sagittal Angle
Sagittal Trunk Angle
Head Height
Angle (deg)
Height (m)
Time (s)
Vicon
L515
L515 - 95% CI
D455
D455 - 95% CI
0
2
4
6
8
140
120
100
80
60
40
20
0
50
−50
20
−20
−40
−60
−80
−100
30
1.6
1.5
1.4
1.3
1.2
1.1
1.0

j)

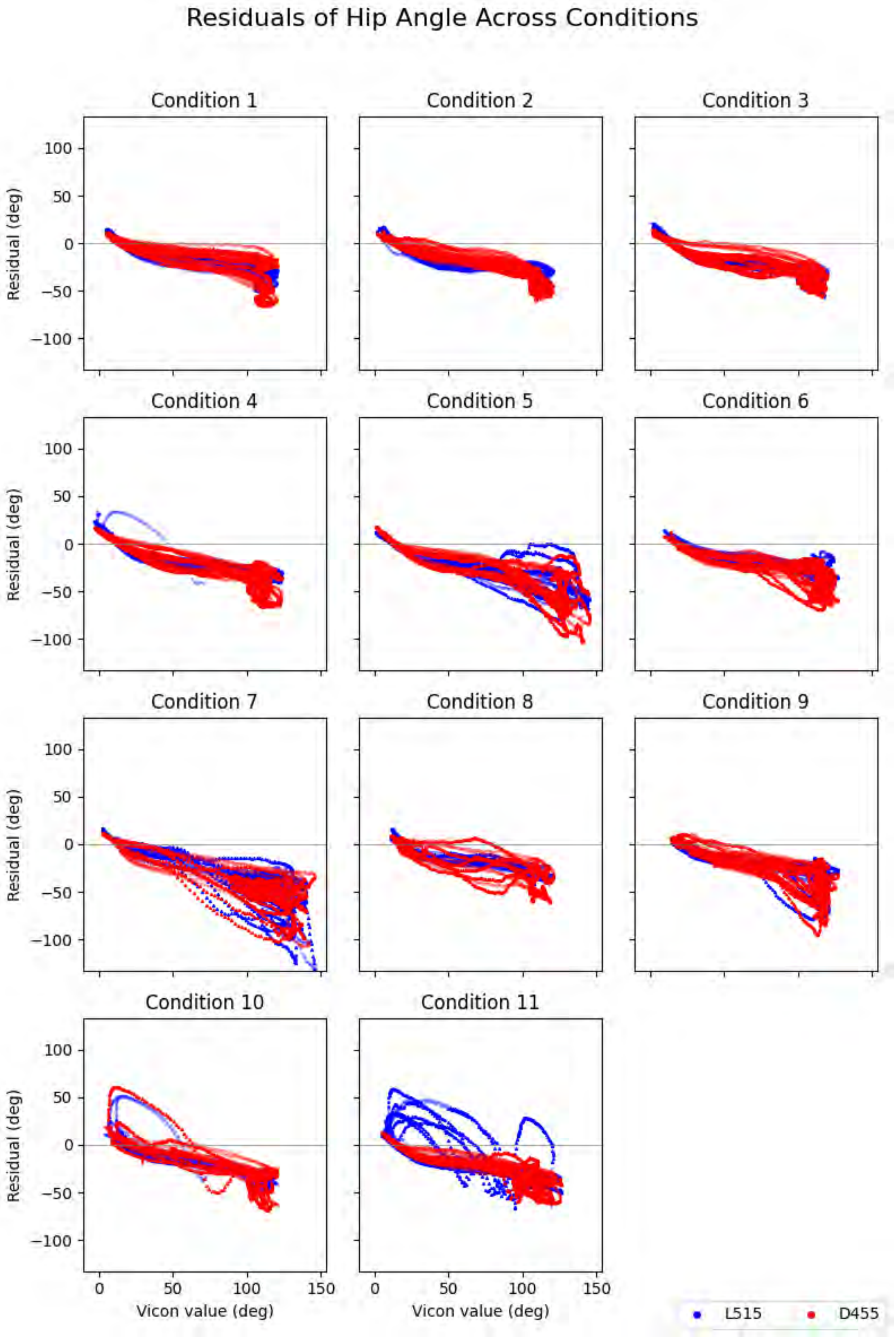


h)

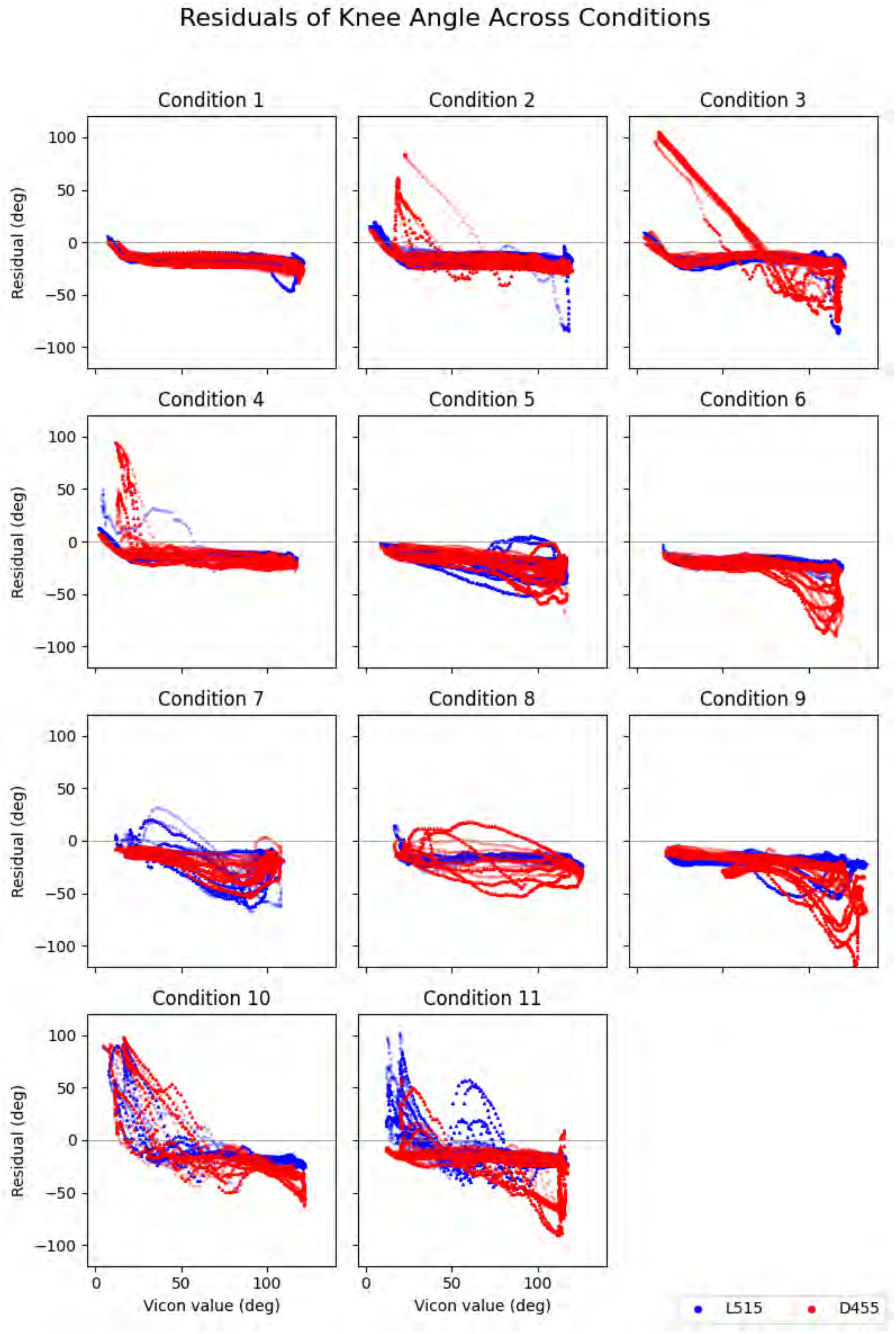
Figure B.2 Offset corrected metrics for the 11 conditions. Conditions 01 to 11 are in order from a) to k).


Residuals of Femur Angle Across Conditions
Condition 1
Condition 2
Condition 3
Condition 4
Condition 5
Condition 6
Condition 7
Condition 8
Condition 9
Condition 10
Condition 11
Residual (deg)
200
100
0
−100
−200
−100
−50
Vicon value (deg)
L515
D455

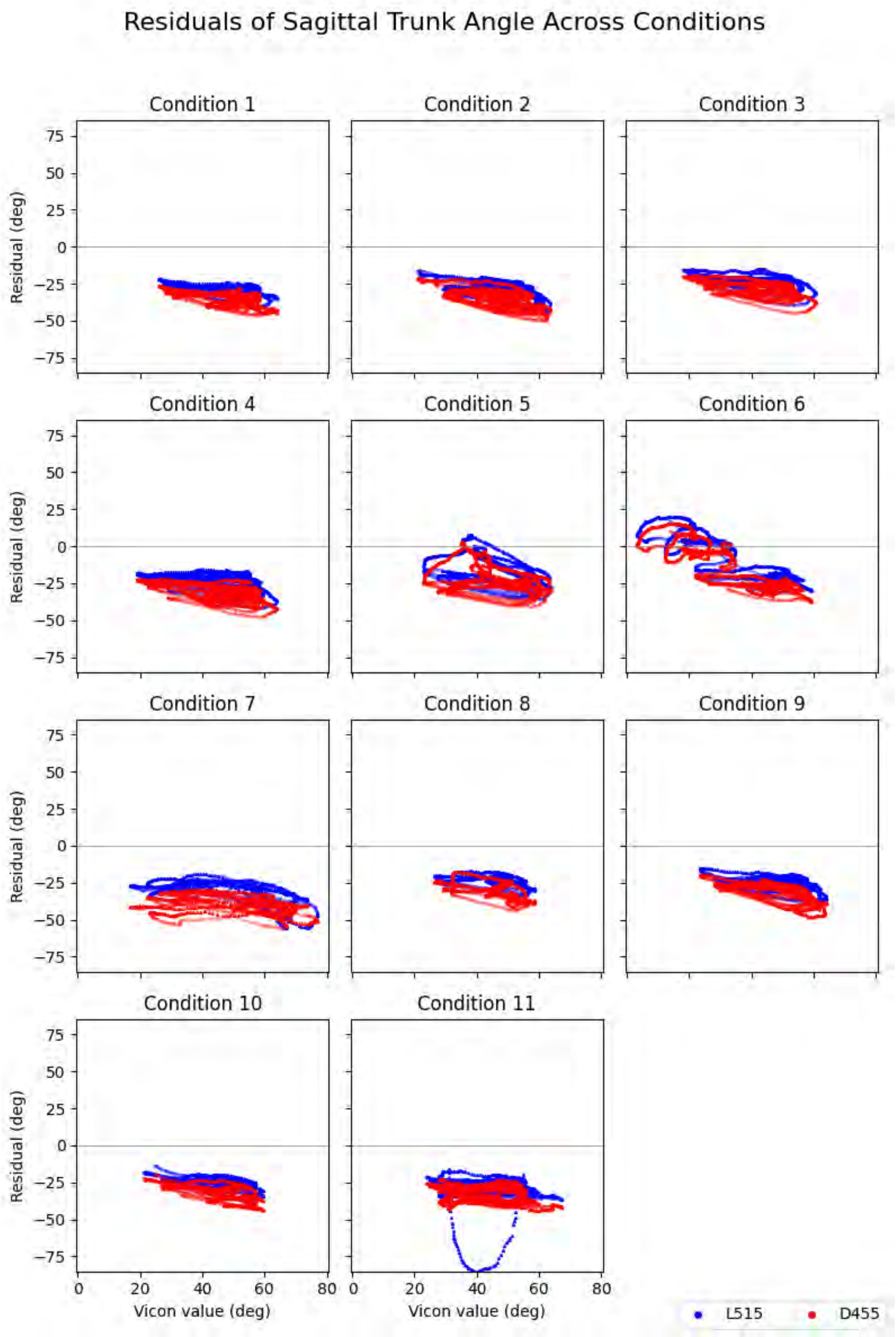
a)


Residuals of Hip Angle Across Conditions
Condition 1
Condition 2
Condition 3
Condition 4
Condition 5
Condition 6
Condition 7
Condition 8
Condition 9
Condition 10
Condition 11
Residual (deg)
100
50
0
−50
−100
0
50
100
150
Vicon value (deg)
L515
D455

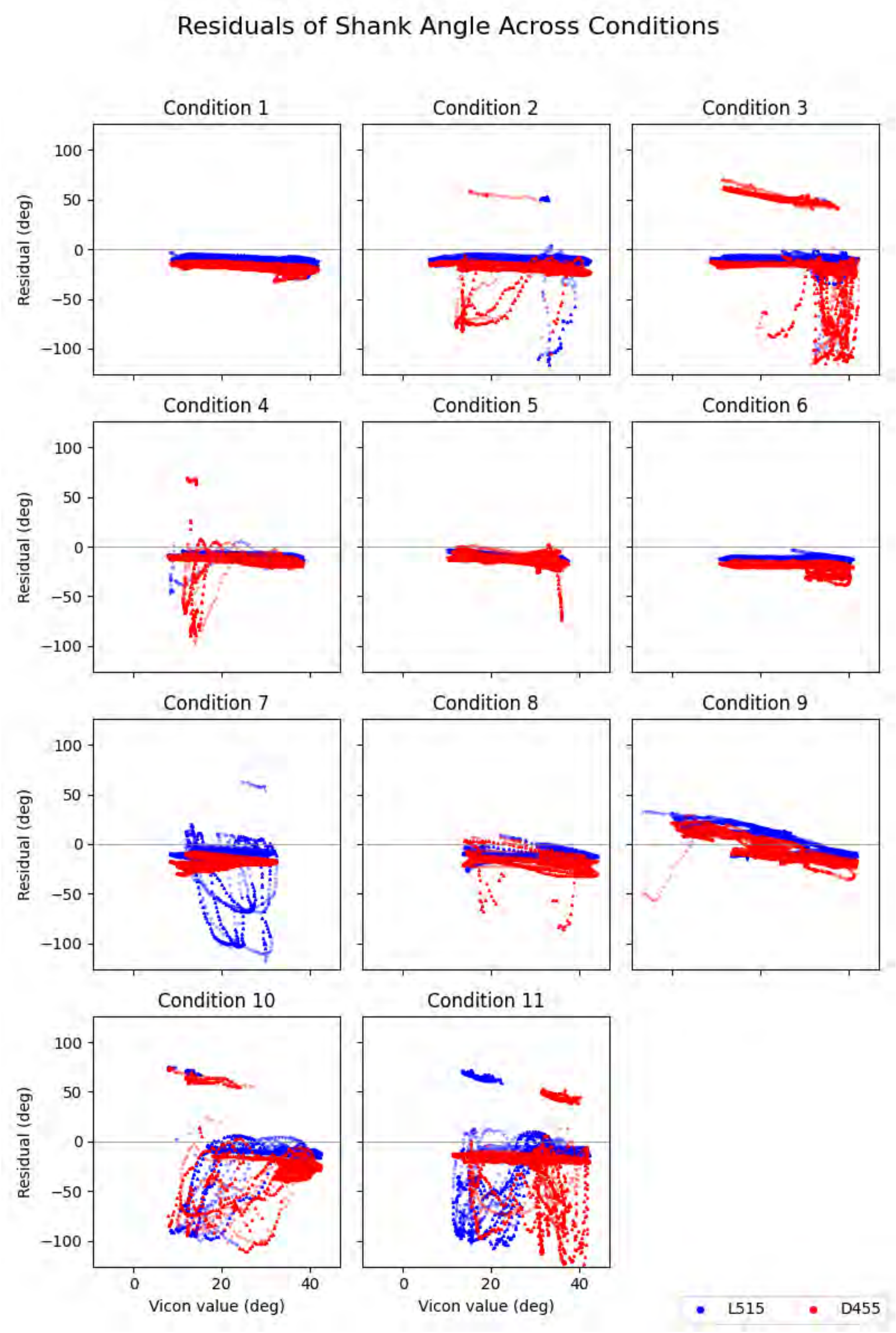
b)


Residuals of Knee Angle Across Conditions
Condition 1
Condition 2
Condition 3
Condition 4
Condition 5
Condition 6
Condition 7
Condition 8
Condition 9
Condition 10
Condition 11
Residual (deg)
100
50
0
−50
−100
Vicon value (deg)
0
50
100
● L515
● D455

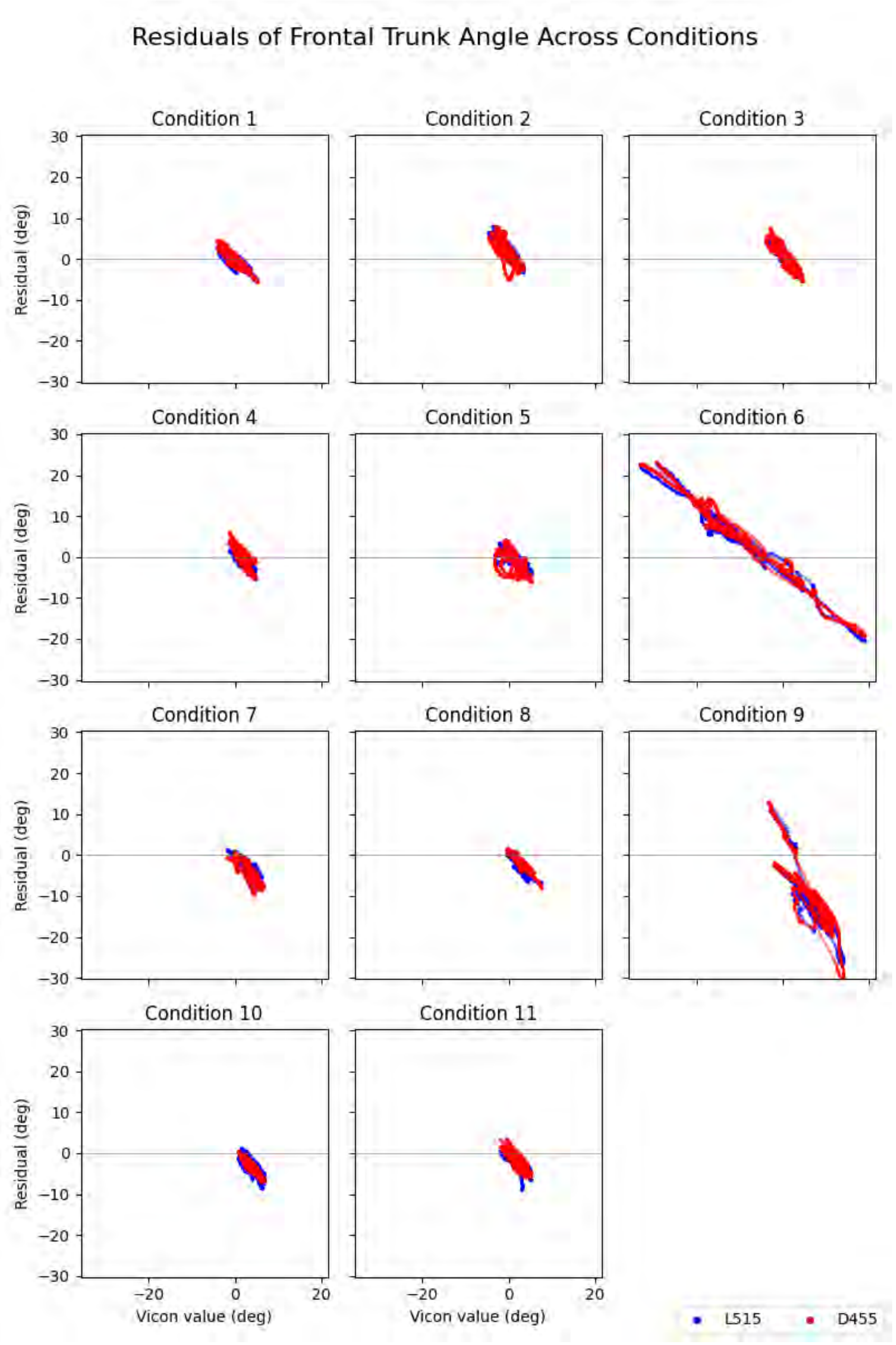
c)


Residuals of Sagittal Trunk Angle Across Conditions
Condition 1
Condition 2
Condition 3
Condition 4
Condition 5
Condition 6
Condition 7
Condition 8
Condition 9
Condition 10
Condition 11
Residual (deg)
75
50
25
0
−25
−50
−75
0
20
40
60
80
Vicon value (deg)
L515
D455

d)

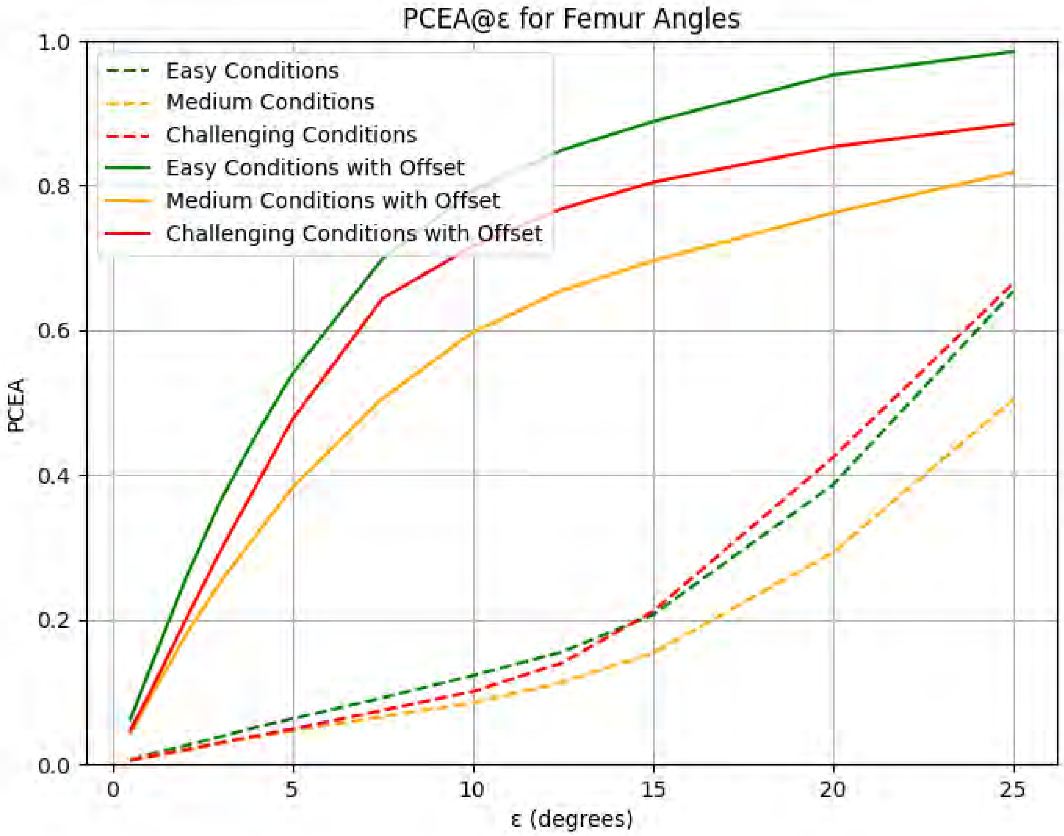

Residuals of Shank Angle Across Conditions
Condition 1
Condition 2
Condition 3
Condition 4
Condition 5
Condition 6
Condition 7
Condition 8
Condition 9
Condition 10
Condition 11
Residual (deg)
100
50
0
−50
−100
0
20
40
Vicon value (deg)
● L515
● D455

e)

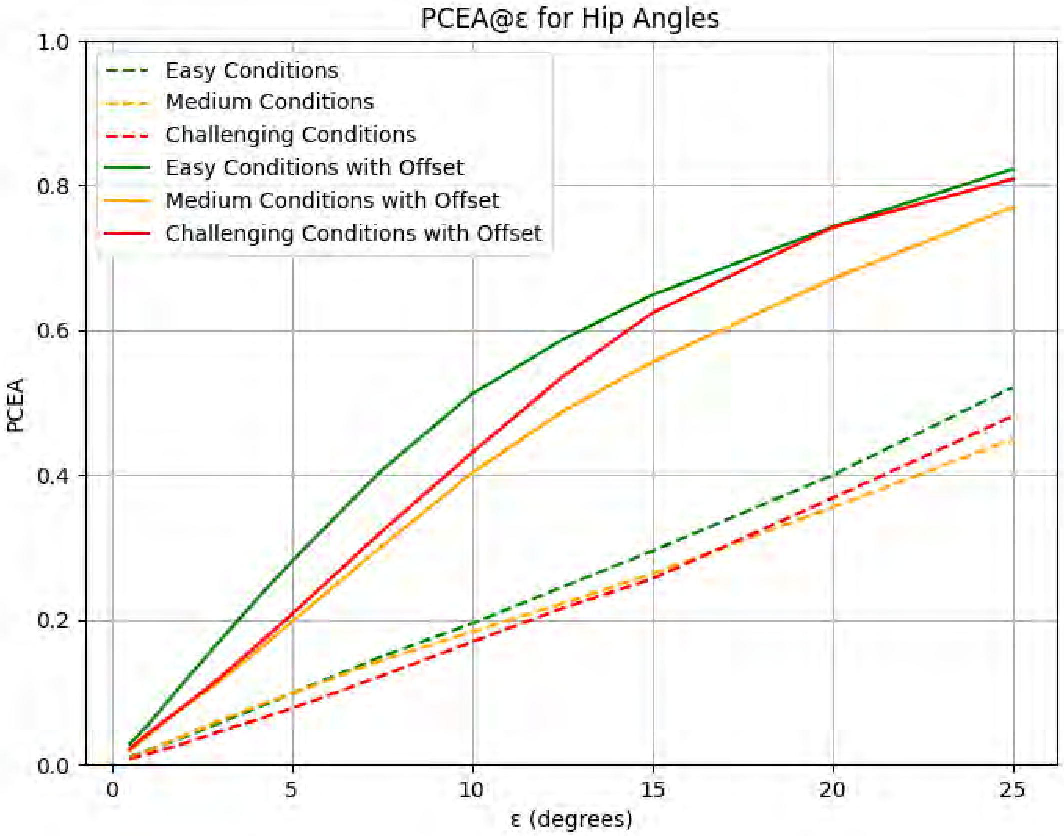


f)

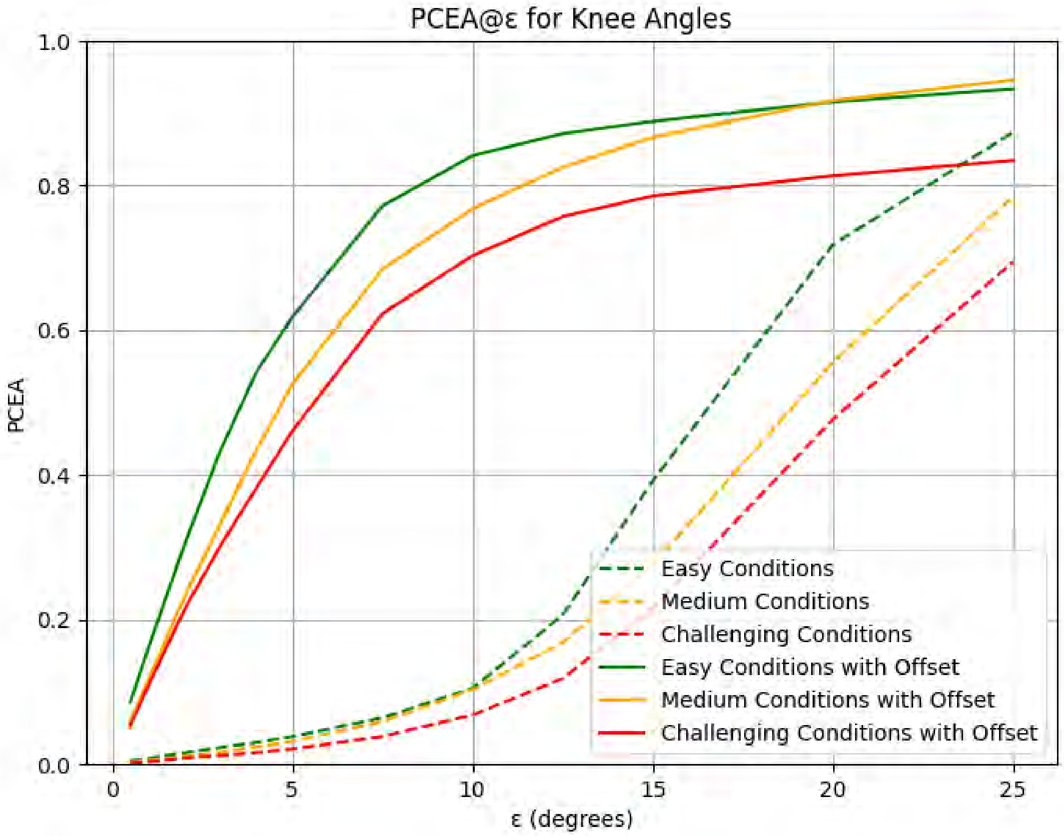
Figure B.3 Residuals for all metrics divided by conditions and camera. Depending on the metric, the residual trend is either flat, angled or noisy.


PCEA@ε for Femur Angles
PCEA
ε (degrees)
0.0
0.2
0.4
0.6
0.8
1.0
0
5
10
15
20
25
Easy Conditions
Medium Conditions
Challenging Conditions
Easy Conditions with Offset
Medium Conditions with Offset
Challenging Conditions with Offset

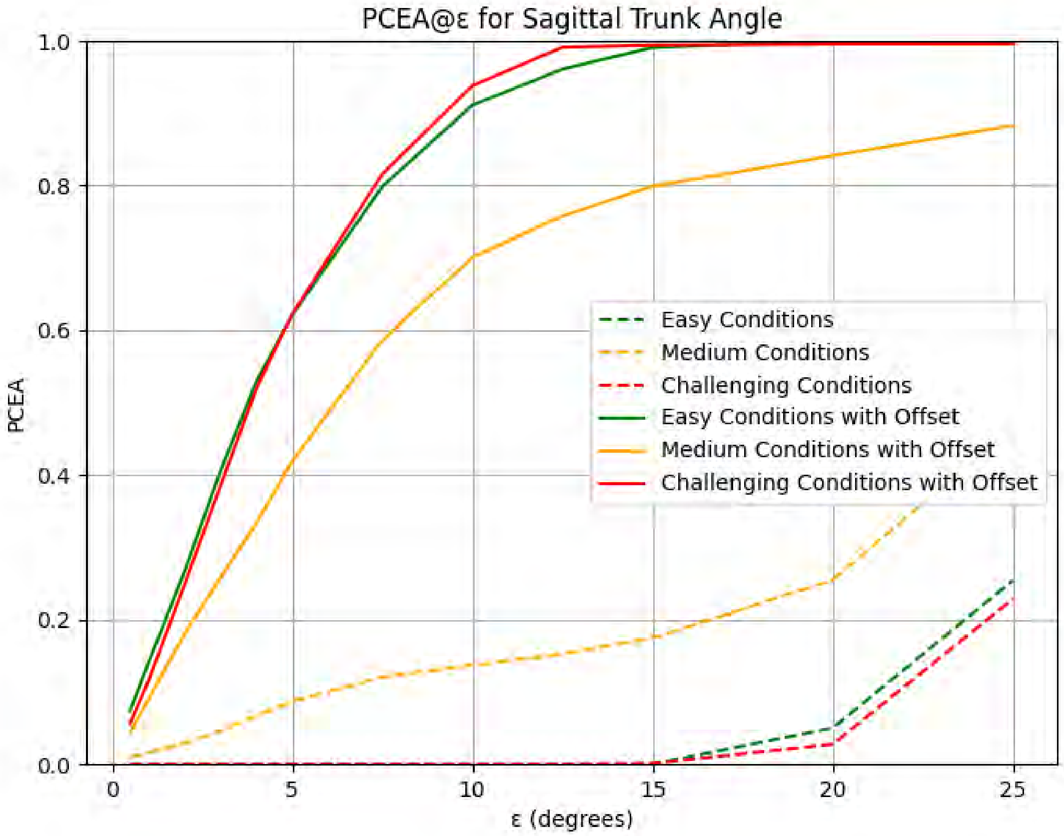
a)


PCEA@ε for Hip Angles
Easy Conditions
Medium Conditions
Challenging Conditions
Easy Conditions with Offset
Medium Conditions with Offset
Challenging Conditions with Offset
PCEA
ε (degrees)
0.0
0.2
0.4
0.6
0.8
1.0
0
5
10
15
20
25

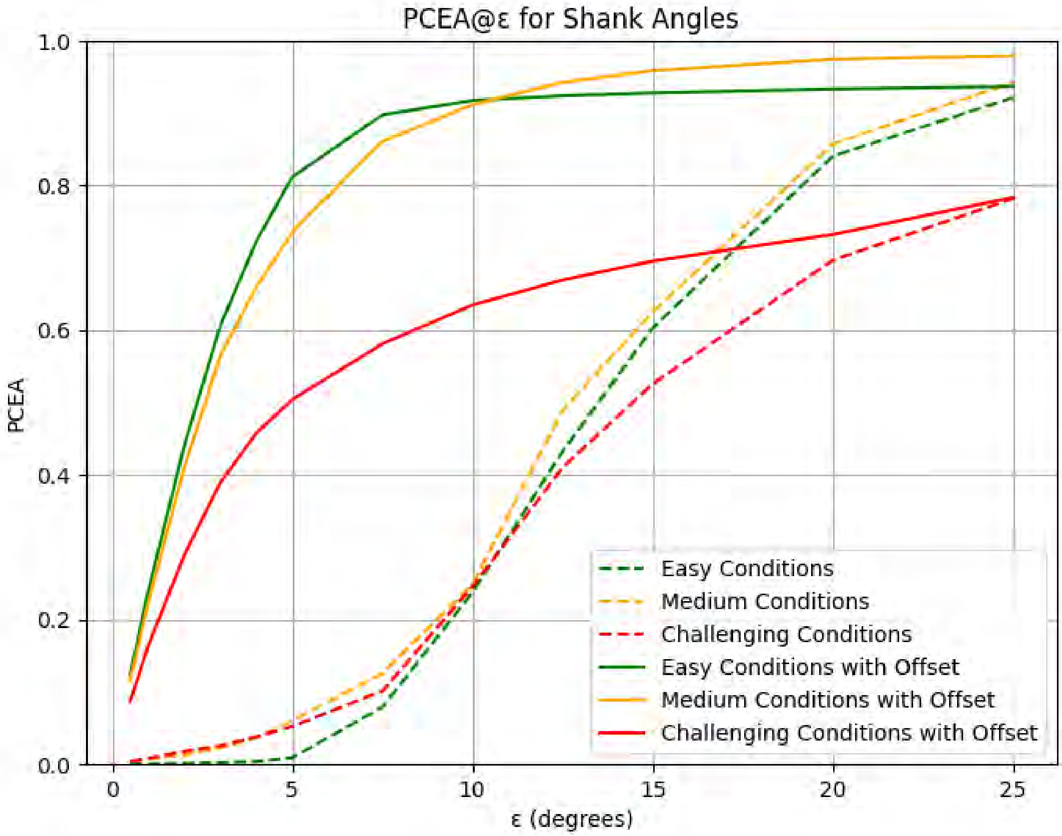
b)


PCEA@ε for Knee Angles
PCEA
ε (degrees)
0.0
0.2
0.4
0.6
0.8
1.0
0
5
10
15
20
25
Easy Conditions
Medium Conditions
Challenging Conditions
Easy Conditions with Offset
Medium Conditions with Offset
Challenging Conditions with Offset

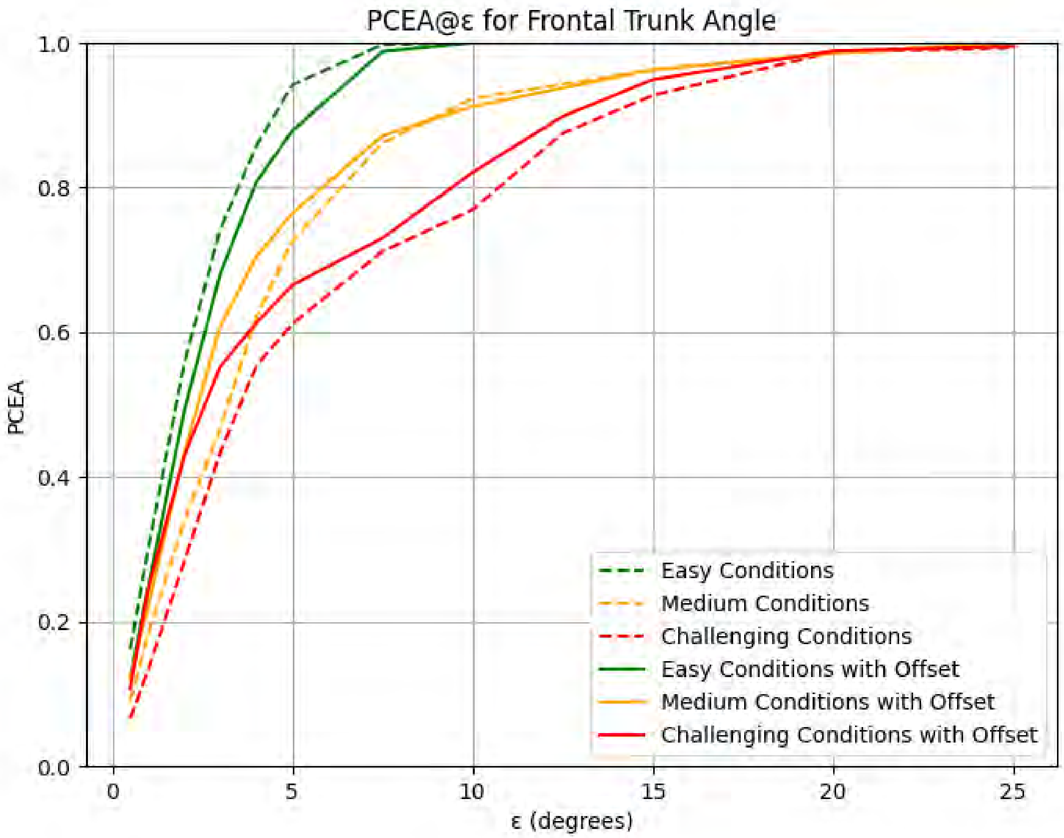
c)


PCEA@ε for Sagittal Trunk Angle
PCEA
ε (degrees)
0.0
0.2
0.4
0.6
0.8
1.0
0
5
10
15
20
25
Easy Conditions
Medium Conditions
Challenging Conditions
Easy Conditions with Offset
Medium Conditions with Offset
Challenging Conditions with Offset

d)


PCEA@ε for Shank Angles
PCEA
ε (degrees)
0.0
0.2
0.4
0.6
0.8
1.0
0
5
10
15
20
25
Easy Conditions
Medium Conditions
Challenging Conditions
Easy Conditions with Offset
Medium Conditions with Offset
Challenging Conditions with Offset

e)



f)

Figure B.4 PCEA@ϵ plots for all metrics before and after correction with the offset.

Table B.1 Velocity metrics per condition and peak type. The positive peak is the average of the maximum positive velocity per repetition for that condition and the negative peak averages negative speed peaks.

Condition	Peak Type	System	Head X (%)	Head Z (%)	Left Knee (%)	Right Knee (%)	Left Hip (%)	Right Hip (%)	Left Femur (%)	Right Femur (%)	Left Shank (%)	Right Shank (%)	Sagittal Trunk (%)
1	Positive	**L515**	-4.4	9.6	3.9	5.4	-10.2	-16.0	-8.0	-3.9	-21.0	-21.1	-13.4
		D455	19.5	11.8	3.9	-2.3	-15.3	-24.5	-5.7	-20.9	-12.4	-27.6	-13.6
	Negative	**L515**	1.2	-0.2	-3.4	-7.5	-14.0	-6.3	-18.9	-15.3	-22.4	-26.3	-32.0
		D455	22.9	-0.5	0.7	-22.8	-9.5	-26.5	-17.9	-18.9	-20.9	-23.0	-30.1
2	Positive	**L515**	1.9	14.2	7.6	16.0	-5.7	-9.2	-4.3	3.3	97.5	-13.1	-10.3
		D455	16.2	18.3	3.6	34.5	-10.7	-26.5	-2.4	2.5	-15.9	357.8	-2.2
	Negative	**L515**	3.1	4.4	6.7	2.8	-15.8	-14.9	-9.9	1.5	91.3	-20.2	-33.9
		D455	19.5	9.0	0.2	24.5	-8.5	-16.9	-10.3	-9.7	-12.8	240.3	-14.8
3	Positive	**L515**	4.3	18.4	52.1	27.8	2.8	-13.5	-16.7	-0.3	232.3	1.3	-2.2
		D455	14.2	14.6	29.1	104.3	-5.0	0.7	-15.0	47.8	4.3	1350.7	-7.8
	Negative	**L515**	4.8	9.8	11.5	17.2	-15.9	-14.3	3.3	3.3	255.7	-7.0	-27.8
		D455	11.4	10.5	8.9	70.3	-11.0	-11.3	1.8	46.7	-2.7	1067.6	-21.2
4	Positive	**L515**	6.7	14.7	20.9	17.4	-3.1	-13.3	10.8	-0.7	120.5	-12.9	-2.4
		D455	20.5	14.9	-3.8	88.6	-16.6	-12.8	-15.0	-5.6	-20.6	918.2	-7.3
	Negative	**L515**	-1.0	6.4	11.0	7.4	-11.8	-9.6	-2.4	-18.7	89.2	-9.6	-32.8
		D455	17.5	7.0	3.6	67.8	-26.3	-25.2	-16.1	-14.5	-13.1	479.4	-6.5
5	Max	**L515**	-17.2	2.4	-4.4	3.2	-20.4	-21.0	-10.2	-2.2	-30.5	-25.4	-8.3
		D455	32.3	22.2	27.3	1.5	2.5	-8.8	8.7	1.9	-23.9	167.6	34.8
	Min	**L515**	-24.0	-9.9	-18.6	-12.8	-15.8	-7.5	-29.8	18.5	-38.0	-33.9	-23.6
		D455	15.4	11.9	20.0	-3.9	5.2	-21.7	8.2	17.8	-10.2	51.2	-6.5
6	Max	**L515**	-4.3	9.4	19.3	13.7	1.3	4.7	-9.0	4.3	-3.1	-15.9	7.4
		D455	-3.8	13.3	15.6	8.8	-16.7	14.6	15.7	0.4	8.2	-4.7	9.4
	Min	**L515**	9.9	42.6	12.6	9.2	-11.3	29.0	0.2	41.0	-7.3	28.6	-5.3

		D455	17.1	29.0	7.1	10.7	8.1	45.0	22.6	36.8	-5.4	39.1	1.2
7	**Max**	**L515**	3.8	12.8	34.6	14.2	-8.3	-4.3	9.7	37.9	797.0	22.1	-17.7
		D455	7.6	1.4	-6.5	-1.0	-40.7	-29.3	139.3	159.6	1.6	52.8	-20.6
	Min	**L515**	7.5	4.5	21.8	-2.7	-18.0	-12.7	28.1	36.2	527.1	-31.0	-0.6
		D455	42.1	-4.0	-21.3	-19.7	-22.2	-21.7	129.8	149.0	-21.5	6.4	-16.4
8	**Max**	**L515**	-4.9	14.8	9.8	14.5	-6.5	-4.5	-13.8	-10.5	-28.5	5.8	-0.7
		D455	103.0	21.5	3.1	46.3	-0.5	-6.2	-12.0	13.1	9.0	475.9	-5.5
	Min	**L515**	-2.1	5.7	1.7	-1.4	-16.4	-10.2	-11.5	-3.8	-30.2	4.1	-25.6
		D455	10.8	3.7	20.2	19.6	-7.0	-28.4	-6.6	-13.2	17.7	368.9	-16.3
9	**Max**	**L515**	7.0	15.9	16.8	18.8	13.3	-2.8	-5.1	-12.5	-43.5	-10.2	-6.3
		D455	21.8	19.2	45.4	10.9	1.2	7.5	9.0	-6.4	-10.4	-28.6	-8.9
	Min	**L515**	1.0	13.4	-7.7	9.1	-4.8	-20.4	17.7	11.0	-36.2	-8.3	-21.6
		D455	10.4	11.6	15.5	0.4	-3.6	-5.1	61.9	21.7	9.4	36.9	-12.0
10	**Max**	**L515**	9.7	11.3	34.4	96.8	0.9	-4.1	23.8	20.2	433.0	898.5	-21.0
		D455	28.7	14.6	65.4	78.9	-5.7	-11.3	45.6	0.8	363.3	812.8	-15.8
	Min	**L515**	3.7	1.4	24.6	81.4	-24.3	-23.7	16.9	-2.8	376.6	566.6	-36.4
		D455	21.9	3.1	58.4	-0.6	-1.2	-19.2	37.4	-9.0	353.8	408.8	-20.8
11	**Max**	**L515**	4.3	11.6	100.4	150.8	31.8	14.0	91.4	98.4	793.0	1164.8	42.3
		D455	11.8	12.4	-1.0	111.1	-17.3	41.7	-19.2	283.8	-25.4	1277.9	-17.1
	Min	**L515**	4.2	15.5	60.5	86.8	48.2	45.9	118.2	103.7	526.6	813.3	41.4
		D455	15.0	7.2	3.1	77.8	-18.9	15.2	-25.5	242.3	-17.7	850.3	-12.1

Table B.2 RMSE Values from the non-corrected (a) and corrected (b) results.

index	**Right Knee**	**Left Knee**	**Right Hip**	**Left Hip**	**Right Shank Sagittal**	**Left Shank Sagittal**	**Right Femur Sagittal**	**Left Femur Sagittal**	**Sagittal Trunk**	**Frontal Trunk**	**All Metrics**
01D	21.75	18.1	37.49	14.53	19.2	17.97	21.84	15.92	35.95	1.94	20.47
01L	15.31	18.23	26.28	23.5	9.23	13.09	23.21	22.8	29.26	1.74	18.27
02D	28.29	17.66	30.53	29.64	32.68	16.21	20.52	19.03	35.9	3.48	23.39
02L	15.67	22.07	27.3	24.65	9.61	20.01	22.51	23.2	29.39	3.6	19.8
03D	53.05	16.62	31.72	20.98	59.69	16.32	25.62	20.29	29.99	2.92	27.72
03L	14.8	22.37	26.81	27.02	9.01	19.8	25.29	25.83	24.03	2.58	19.75
04D	28.54	15.73	38	23.72	31	14.21	25.07	22.58	32.43	2.63	23.39
04L	13.01	17.97	24.89	26.51	9.22	15.76	25.25	26.38	25.33	1.96	18.63
05D	23.76	21.16	50.83	30.52	13.4	15.35	47.36	38.06	26.24	2.97	26.96
05L	20.92	25.03	34.99	38.69	9.73	10.73	43.45	47	21.11	2.07	25.37
06D	22.61	41.73	30.47	33.93	18.44	22.29	22.11	33.55	20.81	10.42	25.64
06L	20.16	21.02	27.64	22.66	12.15	12.14	24.58	23.41	16.16	10.03	18.99
07D	22.58	22.7	50.09	41.15	18.01	16.89	58.46	48.69	42.69	6.24	32.75
07L	19.58	27.33	44.41	46.07	9.99	39.12	44.11	42.84	32.45	4.25	31.02
08D	28.46	24.62	36.48	20.83	32.4	17.59	26.02	21.41	31.38	2.95	24.21
08L	16.24	18.94	23.7	22.36	10.29	11.01	26.58	26.31	24.68	3.5	18.36
09D	22.52	46.56	29.21	36.56	16.07	12.82	23.44	37.06	32	12.46	26.87
09L	19.42	21.58	32.62	24.18	10.68	16.08	27	23.87	27.26	13.02	21.57
10D	40.25	39.93	34.84	26.18	48.4	38.76	19.64	24.42	32.66	3.71	30.88
10L	30.18	28.01	30.12	27.38	35.77	30.09	23.97	24.42	27.01	3.73	26.07
11D	39.63	18.43	34.55	27.94	53.06	16.51	32.14	18.84	34	2.51	27.76
11L	25.3	26.63	32.33	33.14	33.65	34.09	25.75	27.21	29.9	2.79	27.08
All L	19.15	22.65	30.1	28.74	14.49	20.17	28.34	28.48	26.05	4.48	22.26
All D	30.13	25.75	36.75	27.82	31.12	18.63	29.29	27.26	32.19	4.75	26.37
All Files	24.64	24.2	33.42	28.28	22.8	19.4	28.81	27.87	29.12	4.61	24.32

(a)

index	Right Knee	Left Knee	Right Hip	Left Hip	Right Shank Sagittal	Left Shank Sagittal	Right Femur Sagittal	Left Femur Sagittal	Sagittal Trunk	Frontal Trunk	All Metrics
01D	7.12	7.64	21.91	17.43	3.46	4.25	6.3	7.41	6.13	2.16	8.38
01L	4.24	7.19	14.01	14.61	2.67	4.18	6.49	7.33	5.17	1.94	6.78
02D	23.8	10.11	18.35	19.86	24.43	2.87	8.48	8.47	7.2	4.19	12.77
02L	5.49	15.56	15.05	15.5	2.6	15.64	9.89	8.33	6.48	4.58	9.91
03D	56.42	8.13	17.72	16.86	61.73	2.66	13.26	8.99	5.77	3.37	19.49
03L	4.17	13.87	14.43	16.67	2.61	15.2	10.12	9.27	5.14	3.02	9.45
04D	28.43	9.81	23.71	18.83	27.24	4.06	11.26	11.46	6.22	2.75	14.38
04L	4.9	13.19	15.85	19.21	3.08	9.96	10.81	12.21	5.58	1.87	9.67
05D	12.92	12.35	33.25	21.33	11.72	3.81	30.74	22.29	12.56	2.56	16.35
05L	11.27	15.03	21.12	25.41	2.64	3.22	24.04	27.36	12.56	1.87	14.45
06D	6.31	26.77	16.13	21.1	3.02	7.52	10.02	20.59	21.51	10.78	14.38
06L	5.94	7.23	13.97	14.89	2.59	3.01	9.69	10.04	21.24	10.41	9.9
07D	11.86	15.41	32.13	25.97	7.06	5.86	51.23	43.16	13.4	5.41	21.15
07L	10	19.91	27.58	30.39	3.85	35.36	35.96	33.47	9.77	3.27	20.96
08D	19.23	16.46	21.66	17.36	25.06	11.52	12.13	9.49	5.81	2.35	14.11
08L	4.62	6.96	11.12	15.31	3.43	6.42	5.56	5.48	4.22	2.52	6.56
09D	10.1	30.82	20.28	21.8	9.43	22.28	10.61	22.31	5.93	11.51	16.51
09L	6.83	7.73	18.72	11.65	8.58	25.18	9.04	8.79	5.5	11.93	11.39
10D	42.31	34.64	23.05	23.77	45.5	30.2	7.69	16.73	5.84	2.84	23.26
10L	32.13	25.96	19.96	21.78	34.43	26.49	11.14	14.56	4.96	2.77	19.42
11D	30.16	6.59	20.44	16.5	47.96	3.05	27.84	16.88	5.91	2.12	17.75
11L	25.8	23.6	21.99	25.52	31.84	30.62	22.88	25.79	8.46	1.9	21.84
All L	10.49	14.2	17.62	19.18	8.94	15.94	14.15	14.79	8.1	4.19	12.76
All D	22.6	16.25	22.6	20.07	24.24	8.92	17.23	17.07	8.75	4.55	16.23
All Files	16.55	15.23	20.11	19.63	16.59	12.43	15.69	15.93	8.43	4.37	14.49

(b)