

POLITECNICO DI TORINO



Master's degree in Biomedical Engineering
A.y. 2021/2022
Graduation session July 2022

**Analysis of the accuracy and reliability of a
telerehabilitation system using a set of inertial
measurement units on healthy subjects**

Master's degree Thesis

Supervisor:

Prof. Andrea Cereatti

Co-relator:

Dott. Ing. Marco Caruso

Candidate:

Giulia Molè

*A mio padre,
come promesso.*

Abstract

Stroke is one of the main causes of disability and, according to data reported by the Italian national health service, more than 90000 hospitalizations are counted every year. Only the 25% of these patients completely recovers while the 75% survives but with disability. In addition, a stroke costs every year 16 billion to the Italian National Health Service and five billion to families. After a stroke event, the impaired patients need assistance and rehabilitation to restore their independence as much as possible. For mildly disability, rehabilitation is divided in two phases, first in the hospital and then at home where, however, a significant number of patients drop out the therapy for lack of assistance. Recently, telerehabilitation has been playing a fundamental role, representing a potential solution to help patients to continue their therapy. In this context, a home-based neuromotor telerehabilitation system (DoMoMEA) for mild-impaired stroke patients implementing a full-body rehabilitation protocol has been presented. The main goal of the DoMoMEA system was to engage patients during the rehabilitation with motor exergames which can quantitatively monitor their progress over the weeks and at the same time give them motivational feedback. To this end, the real-time joint kinematics was estimated using a network of wearable inertial sensors (IMUs) and a set of biomechanical parameters such as the range of motion (ROM) was extracted. However, the accuracy of the joint kinematics time series and the reliability of the biomechanical parameters could be affected by several factors such as the anatomical calibration procedure, the presence of soft tissues artefacts, and errors due to the IMU orientation reconstruction.

The aim of this work was to assess the accuracy of the joint kinematics estimation as provided by the DoMoMEA system using a stereophotogrammetric system as reference. Moreover, the reliability of the estimated ROM was evaluated in terms of absolute agreement using the intra-class correlation coefficient (ICC) during a test-retest session of nine planar rehabilitation exercises including flex/extension of the elbow, wrist, knee, hip, ankle and trunk, ab/adduction of the shoulder, and the trunk rotation around both the vertical and the antero-posterior axes. Nine healthy subjects (25 ± 2.1 y.o.) were tested and equipped with eight wearable inertial measurement units (IMUs) applied with elastic straps on the body segments of interest and 49 retro-reflective markers (Davis protocol) for the reference joint kinematics. For the first test session, subjects performed 20 repetitions of each exercise at about 1/3, 2/3 and 3/3 of their maximum ROM and the values obtained with both DoMoMea and stereophotogrammetric systems were used to compute the root mean square difference (RMSD) between the time series. During retest subjects performed 20 repetitions of each exercise, at their maximum ROM. The

twenty ROM values were thus averaged to calculate the ICC(3,k). During the tested exercises and the different motion amplitudes, average RMSD amounted to less than 6 deg, which is acceptable according to McGinley (2009), for all joints but for the elbow (13 deg). These larger differences observed at the elbow joint are probably due to a different definition of anatomical upper arm axes for the two systems. The ICC values highlighted a good reliability of the IMU-based measurements for the exercises of the wrist, knee, ankle, and trunk rotation around antero-posterior axis with ICCs between 0.79 and 0.89, but a moderate reliability for the elbow which amounted to 0.64 despite highly repeatable movement within session. A moderate reliability resulted also for the shoulder, hip, and the trunk flex/extension, with ICC between 0.69 and 0.73. For hip and shoulder, the reason of these results could be found in the IMU's elastic strap movement caused by muscle contractions, wobbling of soft tissues, and skin stretching/sliding. However, the trunk rotation around the vertical axis presented a poor reliability that amounted to 0.40 probably because the trunk movement was weakly constrained, and its amplitude was difficult to reproduce between sessions. Similar results were observed also for the reference. Overall, the ICC results highlighted a good agreement between the IMU and reference measurements but for the wrist, hip, and trunk rotation around the vertical axis. The findings of the present work are encouraging for future validation on stroke patients in the validation protocol.

Table of contents

1	Clinical relevance	13
1.1	The stroke event and scale of problem.....	13
1.2	The role of telerehabilitation	14
1.3	Aim of the work	16
2	Introduction	17
2.1	The DoMoMEA project	17
2.1.1	Inertial measurements units (IMUs)	18
2.2	Methods for the joint angle estimation.....	20
2.3	Biomechanical parameters of interest	21
3	Experimental session	25
3.1	Material and methods	25
3.2	Experimental protocol	32
4	Data processing.....	37
4.1	Data pre-processing.....	37
4.2	Parameter extraction and metrics definition.....	40
4.2.1	Accuracy	40
4.2.2	Reliability.....	41
5	Results	50
6	Discussion.....	59
7	Conclusions	61
8	Limitations and future work	62
	Appendix: the analysis of variance (ANOVA).....	63
	References.....	67

Glossary of terms

AA Abduction\adduction

ANOVA Analysis of Variance

FE Flex\Extension

ICC Intraclass Correlation Coefficient

IMU Inertial Measurement Units

KAD Knee Alignment Device

MSC Mean Square for Column

MSE Mean Square of Error

MSR Mean Square for Row

RMSD Root Mean Square Difference

ROM Range of Motion

SP Stereophotogrammetric system

TIA Transient Ischemic Attack

ULM Upper Limb Model

Figures

Figure 1.1: The types of strokes in the brain. Illustration from: <https://www.healthline.com/health/stroke> 14

Figure 1.2 : A serious interface of the telerehabilitation system presented on the left and an exergame on the right..... 15

Figure 2.1 : An overview of DoMoMEA; on the left patient's environment involving IMUs, Minix and a TV, on the top a server where data are collected and, on the right, the medical environment. 18

Figure 2.2: The unalignment between anatomical axes and the IMUs axes..... 20

Figure 2.3 : Example of shoulder angle series on top and below a detail of one repetition where the horizontal line represents execution time and the vertical one is the amplitude of movements, the ROM. 23

Figure 2.4 : The outcome shoulder angle series after 10 repetitions at 3/3 of subject's maximum ROM, on top the angle series with offset and below the series after offset removal. 24

Figure 3.1 : Biolab Research of Politecnico di Torino 26

Figure 3.2 : The force plates and active wand v2. 27

Figure 3.3 : Davis protocol on the left and the Upper Limb Model on the right. 27

Figure 3.4 : The set of DoMoMEA; 7 IMUs, elastic straps, Minix and a TV remote controller. 28

Figure 3.5: Static position 30

Figure 3.6 : The experimental equipment for upper limb exercises. 31

Figure 3.7 : The experimental equipment for trials involving lower limbs. 32

Figure 3.8 : Shoulder exercise (on the left) and wrist exercise (on the right)..... 34

Figure 3.9 : From left to right the elbow flex/extension and the trunk rotation around vertical axis. 35

Figure 3.10: The trunk rotation around antero-posterior axis (on the left) and the trunk flex/extension (on the right)..... 36

Figure 3.11 : Respectively from left to right; knee, hip, and ankle flex/extension..... 36

Figure 4.1: The peaks used to synchronize IMU (on the top) and SP. 37

Figure 4.2: In blue SP signal and in orange SP envelope ($f_s=100$ Hz), the blue lines sign the start and the end of the movement, the black dotted line denotes the end of the first series and the red one signs the end of the second series. 38

Figure 4.3: The plot shows SP and IMUs signals to check the synchronization. 38

Figure 4.4: The misalignment reduction.....	39
Figure 4.5: The figure is an example of validation code output: the count of the repetitions for shoulder exercise. The green line denotes the start of the repetition and the blue one the end of the same rep. The title updates when a new repetition is counted.	40
Figure 4.6: The angle waveforms with offset on the left and after the removal on the right. 41	
Figure 4.7: Flowchart to guide researchers in an appropriate selection of ICC (adapted from Terry K. Koo, 2016).....	45
Figure 5.1: Boxplot of exercises involving upper and lower limbs - small amplitude.	54
Figure 5.2: Boxplot of exercises involving trunk - small amplitude.	55
Figure 5.3: Boxplot of exercises involving upper and lower limbs - medium amplitude.....	56
Figure 5.4: Boxplot of exercises involving trunk - medium amplitude.	56
Figure 5.5: Boxplot of exercises involving upper and lower limbs - large amplitude.	57
Figure 5.6: Boxplot of exercises involving trunk - large amplitude.	57
Figure 5.7: Graphic representation of ICC's difference between two systems.	58
Figure A.1: Some examples of distributions with a different level of discrimination (Source: Psychstat – Missouri State).	65

Tables

Table 1: Overview of total recordings during test session for each subject; two is the number of recordings and three are the movement amplitudes requested.	33
Table 2 : Overview of total recordings during retest session for each subject; two is the number of recordings and three are the movement amplitudes requested.	33
Table 3: μ_i, j = the mean value of each matrix M, i=subject, j= experimental session (1=Test, 2= Retest).	42
Table 4: Data matrix to compute ICC.	42
Table 5: Equivalent ICC forms between Shrout and Fleiss (1979) and McGraw and Wong (1996).	43
Table 6: ICC intervals by Koo et al., 2016.	46
Table 7: Accuracy evaluation; RMSD values for shoulder, wrist, and elbow. S1= first recording at 1/3 ROM max S2= second recording at 1/3 ROM max.	50
Table 8: Accuracy evaluation; RMSD values for knee, hip, and ankle. S1= first recording at 1/3 ROM max S2= second recording at 1/3 ROM max.	50
Table 9: Accuracy evaluation; RMSD values for trunk rotation around vertical and antero-posterior axes and trunk flex/extension. S1= first recording at 1/3 ROM max S2= second recording at 1/3 ROM max.	51
Table 10: Accuracy evaluation; RMSD values for shoulder, wrist, and elbow. M1= first recording at 2/3 ROM max M2= second recording at 2/3 ROM max.	51
Table 11: Accuracy evaluation; RMSD values for knee, hip, and ankle. M1= first recording at 2/3 ROM max M2= second recording at 2/3 ROM max.	51
Table 12: Accuracy evaluation; RMSD values for trunk rotation around vertical and antero-posterior axes and trunk flex/extension. M1= first recording at 2/3 ROM max M2= second recording at 2/3 ROM max.	52
Table 13: Accuracy evaluation; RMSD values for shoulder, wrist, and elbow. L1= first recording at 3/3 ROM max L2= second recording at 3/3 ROM max.	52
Table 14: Accuracy evaluation; RMSD values for knee, hip and ankle. L1= first recording at 3/3 ROM max L2= second recording at 3/3 ROM max.	52
Table 15: Accuracy evaluation; RMSD values for trunk rotation around vertical and antero-posterior axes and trunk flex/extension. L1= first recording at 3/3 ROM max L2= second recording at 3/3 ROM max.	53
Table 16: ICC values for both IMU and SP.	58

Table 17: An overview of ANOVA models.....63

Oral presentation in conference

Part of this thesis work resulted in the writing of a summary for a congress entitled “**Test-retest reliability of inertial-based joint kinematics estimation during rehabilitation exercises in a control group.**”

Authors: G. Molè, M. Caruso, S. Bertuletti, D. Balta, A. Zedda, E. Gusai, S. Spanu, A. Pibiri, M. Monticone, D. Pani, A. Cereatti

University: Politecnico di Torino, Turin, Italy, University of Sassari, Sassari, Italy, University of Cagliari, Cagliari, Italy

1 Clinical relevance

1.1 The stroke event and scale of problem

Stroke is a major cause of neurological morbidity and mortality worldwide. In industrialized countries, stroke constitutes the second or third most common cause of death after cardiovascular disease and cancer and the second most common cause of dementia among older individuals, after Alzheimer's disease [1].

According to a report of 2018 performed by “Osservatorio Ictus Italia” in our country people who had a stroke and survived, with disabling outcomes, were about 940,000 but the phenomenon was constantly growing. This report affirmed that 80% of stroke events was represented by the ischemic stroke which ended often with the death after 30 days from the events for 20-30% of the total number of stroke patients.

In the research conducted by R. A. Grysiewicz [2] the stroke is the rapid development of a focal neurological deficit caused by a disruption of blood supply to the corresponding area of the brain and this causes the death of nerve cells in that area. Consequently, neurological functions controlled by that area (which may involve arm or leg movement, speech, vision, hearing, balance, etc.) are lost. According to data found in the “Ministero della Salute” web site the stroke is most common after 55 years old and the risk doubles at each decade. The stroke definition involves two type of disease the ischemic and the hemorrhagic stroke, but it is important to know that a stroke occurs almost always a transient ischemic attack (TIA). TIA is a focal neurological deficit that lasts less than 24 hours, and it is a predictor of stroke in fact the risk of it occurring is greatest in the first 90 days after the attack. Ministero della Salute” web site reports that about one-third of people who suffered of a TIA is more prone to stroke within five years after TIA.

So, stroke can be classified into ischemic (occlusion of a blood vessel) or hemorrhagic (rupture of a blood vessel). Hemorrhagic stroke includes also intracerebral hemorrhage (bleeding within the brain) and subarachnoid hemorrhage (bleeding between the inner and outer layers of tissue covering the brain within the subarachnoid space). On the other hand, exist some subtypes of ischemic stroke based on the mechanism of injury. These subtypes include atherosclerosis of the great arteries, cardiogenic embolism, small vessel occlusive disease, stroke of other determined cause, and stroke of undetermined cause.

Figure 1.1 shows in a sample image what happens in the brain during different types of strokes.

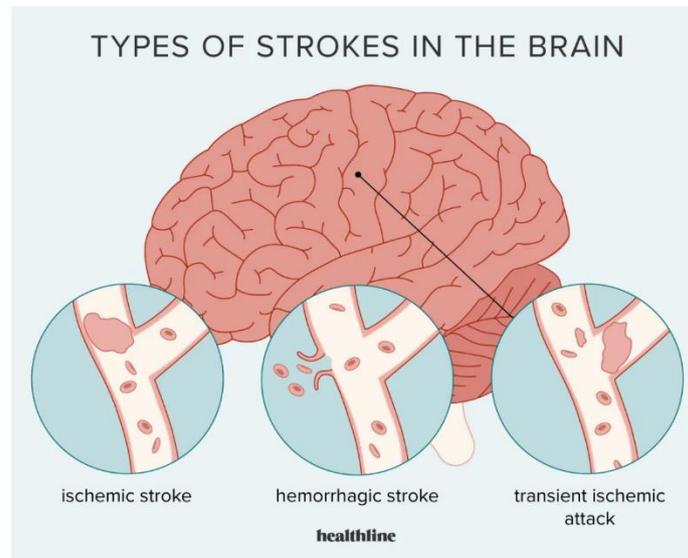


Figure 1.1: The types of strokes in the brain. Illustration from: <https://www.healthline.com/health/stroke>

1.2 The role of telerehabilitation

The types of strokes are different and also the patients survived. Several of them survive but with disability and need assistance and rehabilitation. The therapy is very important to help stroke survivors to restore as much motor control as possible and to return to their normal and independent lives. Typically, for mildly and moderately impaired patients the rehabilitation, is delivered in two separates but equally important phases, first in the hospital and then at home, after discharge [3]. In the first phase patients are assisted by clinicians, but at home they are given only instructions to try to perform exercises autonomously without supervisors. Thus, leading several patients to drop out the therapy for the lack of desire, ability, and assistance. In addition, it is just the absence of supervision during exercises that causes uncorrected compensatory movements, which should be performed properly to reach the desired goal instead they are often unnatural and incorrect. Taking into account of this tendency, the telerehabilitation could be a solution to help therapists to monitor patients at distance so guide them in a correct execution of the exercises but also optimize timing, intensity, and duration of therapy.

Telerehabilitation takes care of the delivery of medical rehabilitation services and the support of independent living, using low-cost communications technologies. Telerehabilitation offers several advantages to patients' environment because it is facilitated to continue the therapy and be monitored also without a real time intervention of a specialist. Particularly the motor telerehabilitation is closely dependent on the modern wearable sensors that provide the acquisition of movement data. The presence of sensors, e.g., inertial measurement units (IMU) for the

acquisition of movement cinematics is fundamental in a system for motor telerehabilitation because it allows real-time feedback to the patient on the quality of the movement performed. However, this part of telemedicine involves clinical disciplines such as speech-language pathology, physical therapy, and neuropsychology rehabilitation and about patients especially those post stroke, brain or spinal cord injuries, amputation and orthopedic or developmental impairments [4]. In this work the home based telerehabilitation system, DoMoMEA wants to offer to post-stroke patients a telerehabilitation service based on exergame to make the therapy more interesting. To reach this goal patients can obtain motivational feedback in real-time after each exercise and then funny graphic interfaces were thought like shown in the Figure 1.2 (taken from *Zedda et al., 2020*). This system is based on a network of wearable sensors, IMUs, cited before and implements a full-body rehabilitation protocol that consists in 15 exercises. It is fundamental that during the development of these system users (e.g., patients, clinicians, medical staff) are taken into account [5]. The technology used in the system in fact must be user-friendly and suitable for all ages users, even for those unfamiliar with technology or with some cognitive deficit. In [4], *Brennan et al.*, also reported information about the increasing of the rehabilitation demand, due to the aging population and the real improvement of the patients who received the therapy. But the growth of the demand leded to physician shortages in fact rehabilitation providers were turning to telerehabilitation as a means to improve quality of care and reduce costs.

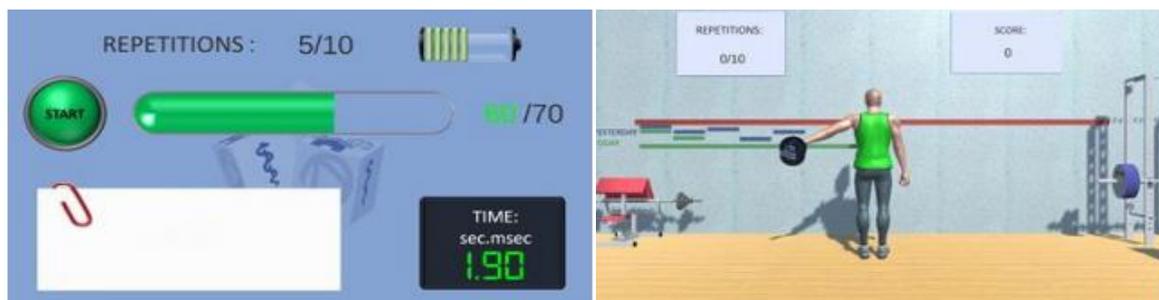


Figure 1.2 : A serious interface of the telerehabilitation system presented on the left and an exergame on the right.

1.3 Aim of the work

The aim of this thesis was to assess accuracy firstly and then the test-retest reliability of tele-rehabilitation system for post-stroke patients: the DoMoMEA system based on wearable inertial measurement units. This system is explained in more detail in the following paragraph. For validation purposes the biomechanical parameter i.e., the joint range of motion (ROM) and the angle time series extracted using the IMUs were compared to those provided with a gold standard for movement analysis research, the stereophotogrammetric system (SP). The experimental protocol consisted in a set of nine planar rehabilitation exercises performed by healthy subjects. The accuracy was assessed in particular by comparing the kinematics extracted during execution of each exercise through the computation of root mean square difference (RMSD) between the angle time series estimated by the SP and the DoMoMEA system. Then test-retest reliability was evaluated by extracting the ROM that was employed to compute the intraclass correlation coefficient (ICC). The latter is the most common metric to assess the capacity of IMUs to give the same results during two different acquisition sessions organized in two different day a week apart. This is important to test IMUs and make sure that they are not affected to errors due to the relocation on body segments of interest.

2 Introduction

2.1 The DoMoMEA project

DoMoMEA which means ‘my home’ in Sardinian language [6] is a home-based neuromotor telerehabilitation system for mild-impaired stroke patients. Nowadays, DoMoMEA represents the only portable rehabilitative system that specifically focuses on post stroke full-body recovery by using wearable and inertial technologies in a game-based training context without assistance. It is a customized and dynamic intervention telemonitored by the clinicians [3]. Real-time joint kinematics is estimated using a network of IMUs and a set of biomechanical parameters, among which the range of motion, are extracted to monitor patients’ progress overtime. Its principal aim is to bring the motor rehabilitation for stroke patients at home even if, this type of rehabilitation protocol is difficult to transfer at home without assistance due to the peculiarities of the impairment and the kind of exercises to be performed. For this reason, the DoMoMEA system provides an architecture which permits to follow patients at a distance give them real-time scores and feedback through home patient’s TV.

In the Figure 2.1 taken from *Zedda et al., 2020* the architecture is reported which is divided into three parts [6]:

- 1) The DoMo system on the left that is composed by all instrumentations at patient’s home: seven IMUs connected via Bluetooth, an Android TV-box, with a simple HDMI connection to patient’s TV and Wi-fi connectivity, which enables the use of a large screen for exergames.
- 2) A server, on the top of the figure, to connect both patient and clinicians and to collect data which will be consulted by specialists.
- 3) On the right the medical environment and the web application where patients can find their own personal data and clinicians can modify the features of each patient’s therapy.

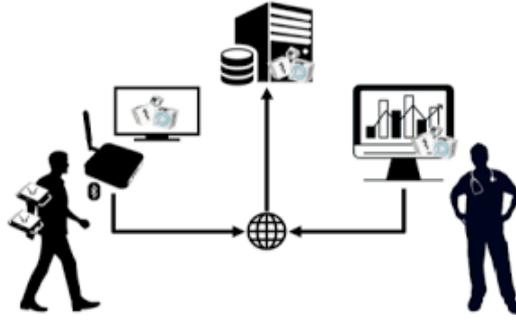


Figure 2.1 : An overview of DoMoMEA; on the left patient's environment involving IMUs, Minix and a TV, on the top a server where data are collected and, on the right, the medical environment.

2.1.1 Inertial measurements units (IMUs)

Traditionally, the human movement is captured by motion capture systems like the stereophotogrammetric system. Although the SP was used as a reference because it is considered the gold standard, but it has some disadvantages because it requires special equipment, is time consuming [7], dependent on patient compliance, and with it is not possible to perform experiments out of laboratory. It is known also that this system is costly, require fixed cameras in a controlled environment, and suffer from marker occlusion. For this reason, an emerging technology with a growing number of potential applications in human movement analysis is in continuous development. These are low-cost wearable inertial sensors which usually containing accelerometers, gyroscopes, and magnetometers with the objective to provide an alternative to overcome the limitations of motion capture systems [8].

Magneto–inertial sensing offers several advantages, but the main is the capability to provide a continuous description of the subject motor performance in his/her specific daily life [9]. In addition, wearable inertial sensors are characterized by low-power consumption and miniaturization.

These systems are composed, typically, by an array of sensors: a three-axial accelerometer, gyroscopes, and magnetometer:

- Magnetometer allows to obtain some information about orientation on horizontal plane, but the last one is not used in this thesis work because of several ferromagnetic disturbances around patient's environment. For this reason, an indetermination on horizontal plane is presented and to solve this problem is good starting from known sensors positions.

However, its output calculation reported in (2.1.1). The \mathbf{h}_{hi} is the hard-iron errors, this term represents an offset due to the permanent magnetization of the ferromagnetic materials which move in solidarity with the magnetometer. In addition, \mathbf{b}_m and \mathbf{w}_m respectively the bias and the white Gaussian noise corrupt the magnetometer output.

$$\mathbf{h} = (\mathbf{h}_{heart} - \mathbf{h}_{hi}) + \mathbf{b}_m + \mathbf{w}_m \quad (2.1.1)$$

- Accelerometer measures specific forces (\mathbf{a}), the vector difference between the acceleration of the body (\mathbf{a}_{body}) and the gravity acceleration (\mathbf{g}), along three axes and allows to derive an angle of inclination reliable in statics with respect to the horizontal. Its main disadvantage is that it doesn't carry any information of orientation in the horizontal plane. A single axis accelerometer can be modelled as a spring-mass model which can move only along the spring's axis [10]. The formula of the accelerometer output reported in (2.1.2) where \mathbf{b}_a is the bias error, and \mathbf{w}_a is the white Gaussian noise:

$$\mathbf{a} = (\mathbf{a}_{body} - \mathbf{g}) + \mathbf{b}_a + \mathbf{w}_a \quad (2.1.2)$$

It is important to note that this system can measure not only dynamic but also static accelerations, such as for example the Earth's gravity. To measure acceleration in all three spatial dimensions three uniaxial accelerometers are mounted perpendicular to each other to form a 3D (triaxial) accelerometer. In static condition the first term of expression is null, so the only sensed acceleration is the gravity. For this reason, is now possible to obtain desired information i.e., the initial inclination of unit with trigonometric formulas. In dynamic instead the terms of expression (2.1.2) are both different to zero in fact the only output acceleration can't distinguish them. With gyroscope will be possible to get to missing information.

- Gyroscope essentially consists of a toroid-shaped rotor that rotates around its axis, when the rotor is rotating its axis tends to keep parallel to itself and resist any attempt to change its orientation. This sensor measures angular velocities along its axes, the main equation of this sensor reported below in (2.1.3) is similar to that of the accelerometer. The \mathbf{b}_g is in fact the gyroscope bias which will be computed and then subtracted to the sensor's readings. At the end \mathbf{w}_g represents the white Gaussian noise.

$$\mathbf{w} = (\mathbf{w}_{body} - \mathbf{g}) + \mathbf{b}_g + \mathbf{w}_g \quad (2.1.3)$$

2.2 Methods for the joint angle estimation

The joints are on basis of human movement analysis which aim is to give quantitative information about the mechanics of the musculo-skeletal system during the execution of a motor task. In particular, the information to be obtained concerns the movement of the whole-body center of mass; the relative movement between adjacent bones [11].

The joint kinematics involves the reconstruction of the relative orientation of the reference systems fixed with the bone segment under examination. This means to calculate temporal trend of six scalar quantities: three relatives to orientation and three relatives to the position. In this section the focus is the explanation of the principles adopted to estimate the joint angle time-series starting from the orientation of the two IMUs attached to the proximal and distal segments of the joint under analysis. The two segments are connected by a spherical joint which allows three degrees of freedom, and the joint kinematics is defined as the relative orientation between the two anatomical axes. One of the first problems in estimation of joint kinematics using IMUs is the unalignment between technical axes of each unit ($x_{Lp}, y_{Lp}, x_{Ld}, y_{Ld}$) and the relevant anatomical axes ($x_{Ap}, y_{Ap}, x_{Ad}, y_{Ad}$) as shown in the Figure 2.2 below.

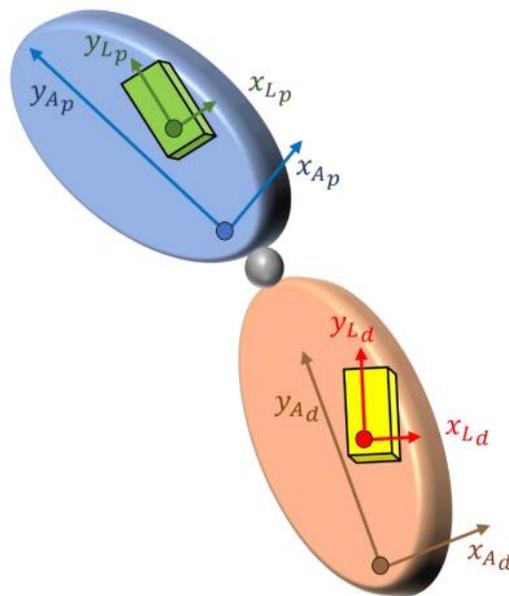


Figure 2.2: The unalignment between anatomical axes and the IMUs axes.

However, it is necessary to estimate the unit orientation which is an output of the device. Although all three sensors in each inertial could be used, individually, to obtain information about the orientation of the rigid body where IMUs are fixed, but this variable is not directly measured by any of them. For this reason, it can be estimated by sensor fusion algorithms which exploit the complementary properties of sensors and address the different sources of error effecting the sensors (like gyroscope bias drift, inertial acceleration, magnetic field distortion) [12].

Sensor fusion algorithms only provide the orientation of the IMU coordinate system, with respect to the global coordinate system which is defined by the directions of the Earth's gravity and magnetic North and in general not aligned with the coordinate system of the body-segment where the IMU is attached (BCS). Therefore, a "sensor-to-segment- alignment" procedure must be implemented to determine the relative orientation between BCS and IMU coordinate system [13].

In this work inertial measurement units consisting of a three axial accelerometer and a three-axial gyroscope, approximately mounted in one of principal body. In theory, a calibrated IMU measures 3D angular velocity and 3D acceleration and gravity with respect to the sensor housing. Given an initial position and orientation, ideally these signals would contain sufficient information to derive the IMU kinematics completely. The orientation can be obtained using a known initial orientation and the change in orientation that can be obtained using gyroscopes [14].

2.3 Biomechanical parameters of interest

Biomechanical parameter, i.e., the joint ROM extracted from the time waveform of angles was employed to quantify patients' progresses. These variables were obtained with IMU-based system and compared with a marker-based system i.e., the SP used as a reference for the validation of the telerehabilitation system. The validation was computed with extraction of ROM, the time of execution and the calculation of RMSD such as an error measure.

Nowadays clinical measurements of ROM have an important impact to choose the correct therapy procedure. For this reason, it is fundamental to interpretate the results correctly because they can have a substantial impact on the development of the scientific basis of rehabilitation inventions [15].

Generally, this parameter is measured before and after therapy treatment to estimate any substantial differences and its measures can be performed on specific joints, and if patient's motion is limited, the therapist can determine if the cause is muscle tightness or pain or tightness of ligaments or tendons [16].

The ROM is defined as the maximum possible amplitude of movement within the limits physiologically imposed by the joints, tendon, and ligament. So, it is an index of joint flexibility, and it is usually measured in degrees. In particular, when a subject performs a movement which involves specific body segments the ROM is the difference between the maximum and the minimum angle reached during the exercise. This quantity has been calculated for both systems and for each joint involved in the study. For example, for the shoulder the ROM computation for SP and IMU is shown in the expressions below:

$$ROM = \max (AngIMU) - \min (AngIMU) \quad (2.3.1)$$

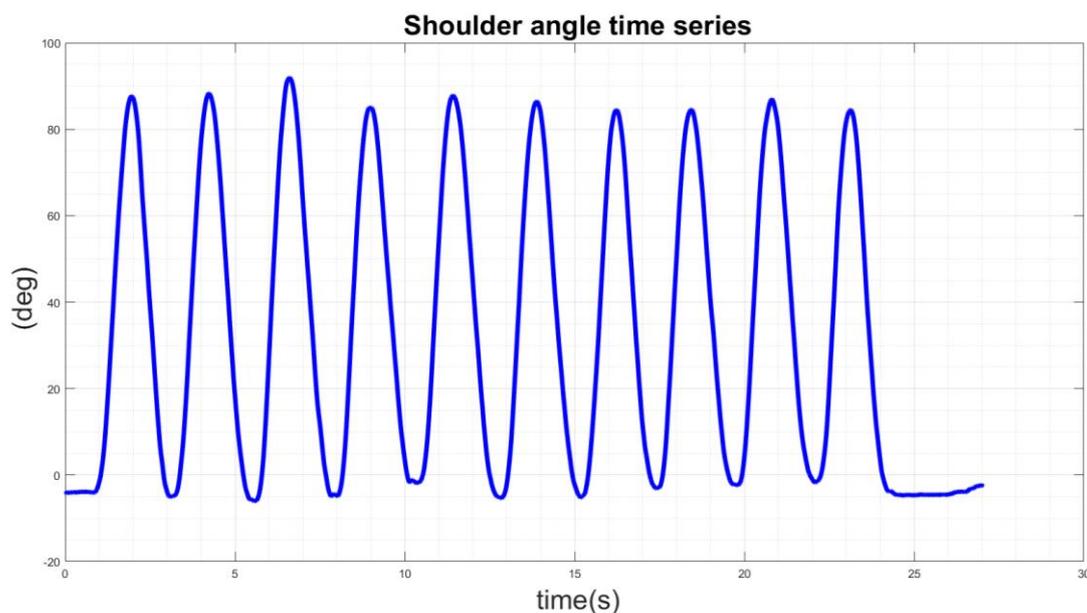
$$ROM = \max (AngSP) - \min (AngSP) \quad (2.3.2)$$

Where ‘AngIMU’ is the angle time series extracted from sensors and ‘AngSP’ is the same angle time series extracted from the reference system. The execution time was also computed and the equations (2.3.3) and (2.3.4) suggest its calculation. The computation of execution time provides the difference between the point corresponding to the end of repetition (endRep) and the point corresponding with the start of repetition (startRep).

$$IMU \text{ execution time} = endRep - startRep \quad (2.3.3)$$

$$SP \text{ execution time} = endRip - startRip \quad (2.3.4)$$

Where ‘*IMU execution time*’ is the time spent to perform the exercise with IMU and ‘*SP execution time*’ is the same quantity obtained with SP.



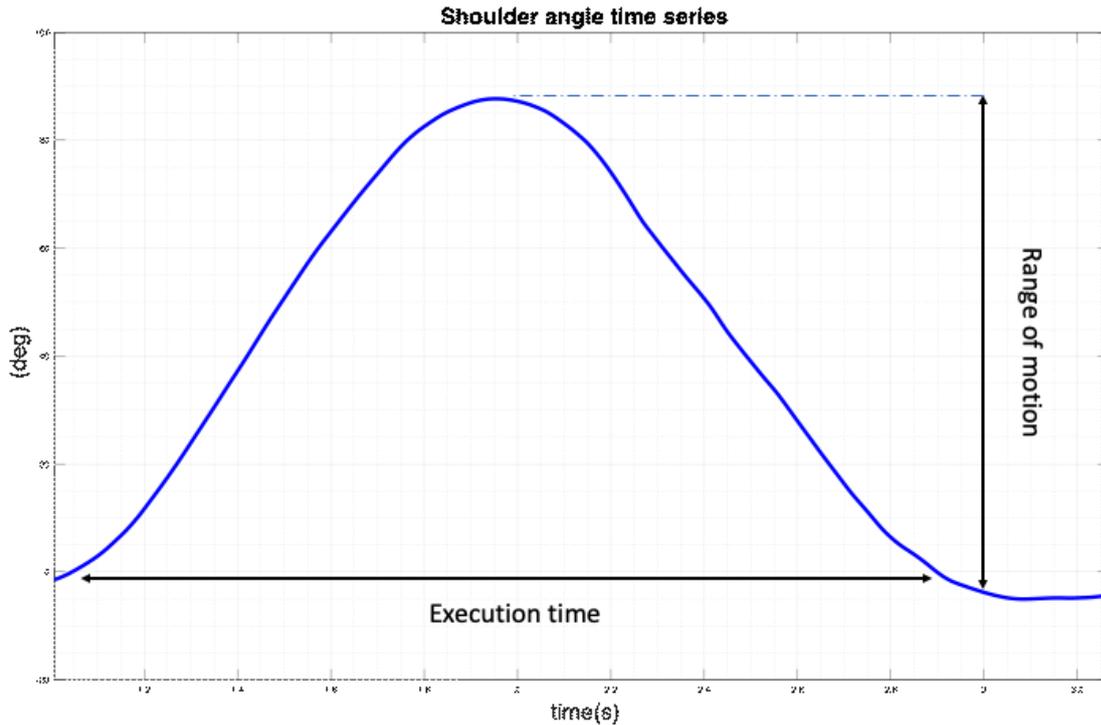


Figure 2.3 : Example of shoulder angle series on top and below a detail of one repetition where the horizontal line represents execution time and the vertical one is the amplitude of movements, the ROM.

The last parameter used for validation is RMSD, i.e., rms of difference between angular curves extracted with IMU and SP system, but after removal of eventual offset, the mean of all angle temporal series.

For each joint the calculation was:

$$RMSD = rms((AngIMU - mean(AngIMU)) - (AngSP - mean(AngSP))) \quad (2.3.5)$$

The mean value, offset, had to be removed due to a different definition of anatomical reference system from the two systems as explained in [17].

RMSD value was an error evaluation between two system, so it was important to validate IMU-based system. McGinley suggested that errors more than 5 deg should raise concern and may be large enough to mislead clinical interpretation [18]. His systematic review was demonstrated that most studies providing estimates of data error reported values of less than 5 deg, except for hip and knee rotation.

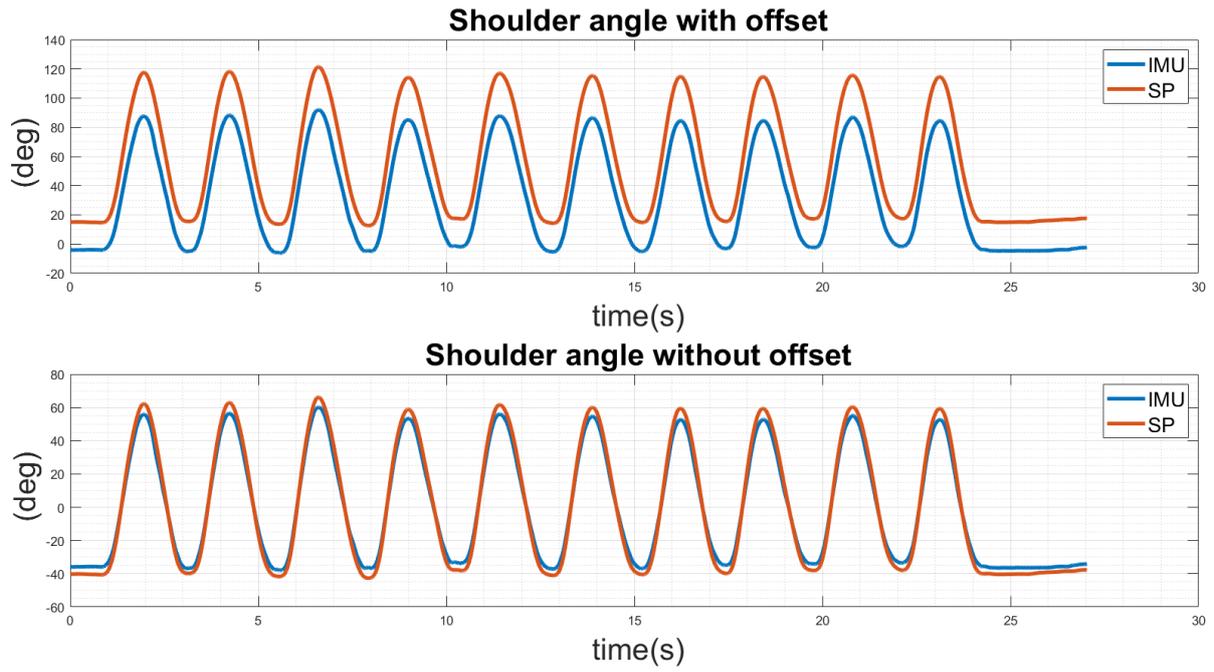


Figure 2.4 : The outcome shoulder angle series after 10 repetitions at 3/3 of subject's maximum ROM, on top the angle series with offset and below the series after offset removal.

3 Experimental session

3.1 Material and methods

The experiments of this work were conducted in the Biolab of Politecnico di Torino showed in Figure 3.1, appropriately equipped with the instrumentations to perform tests of movement analysis.

Two systems were involved:

- Stereophotogrammetric system as the reference
- IMU-based system to validate

The SP required the use of

1. 12 Vicon cameras: infrared cameras, because IR light reduces the artefacts produced by the natural light since the crosstalk is decreased.
2. 3 cameras RGB to record video of experiments.
3. An active wand for calibration of system: a rigid structure (Figure 3.2) on which are mounted marker with a known geometric configuration.
4. 49 passive markers: spherical balls coated with a retroreflective material to reflect IR light which do not require feeding and that have been applied on the subjects according to Davis protocol [19]. This marker set was necessary to use two different models: the Plug in Gait of Nexus + upper limb model (ULM) of GPEM (GPEM s.r.l, Pescara, Italy, <https://www.gpem.net/>) and the correct placement of marker are in Figure 3.3.
5. A force plate AMTI (Advanced Mechanical Technology, Inc Watertown, USA) for the synchronization between two systems.
6. Nexus software (v. 2.12): for extraction of files containing angles and forces.



Figure 3.1 : Biolab Research of Politecnico di Torino

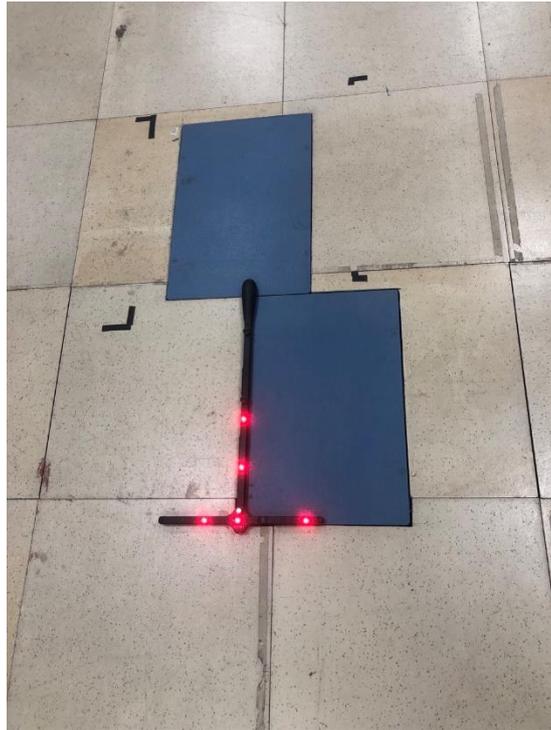


Figure 3.2 : The force plates and active wand v2.

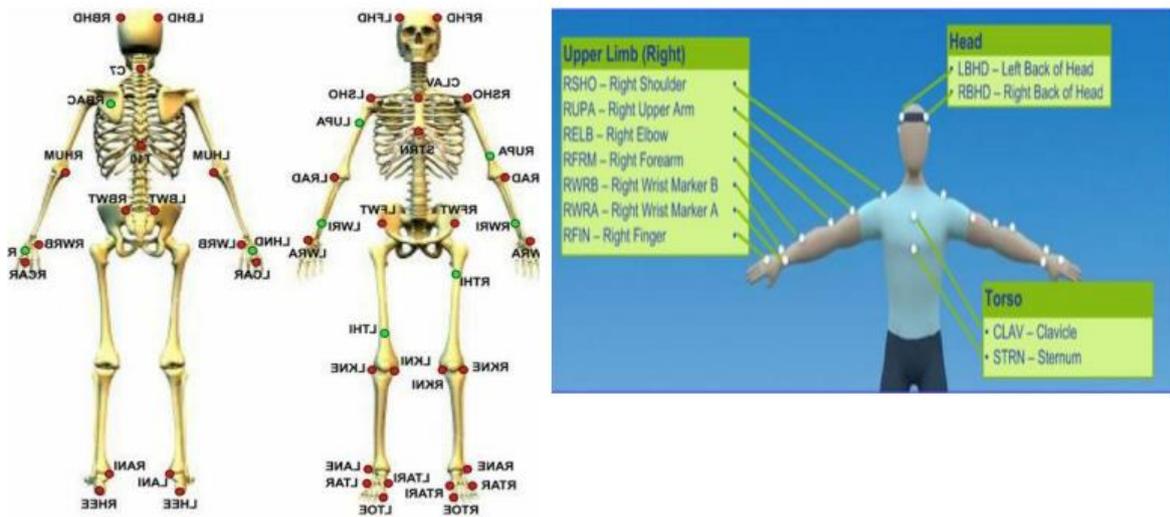


Figure 3.3 : Davis protocol on the left and the Upper Limb Model on the right.

The IMU-based system was composed by

1. Seven IMUs: these sensors were attached on the bands provided; the top of the cases of the IMU was marked with seven different colours to distinguish the body segment on which it should be applied more easily.
2. Elastic straps: the bands were made up of two substrates: a neoprene's substrate and another one of Velcro-plush for the sensors that must be keep firmly in a specific position during exercise.

The complete set is shown in Figure 3.4.



Figure 3.4 : The set of DoMoMEA; 7 IMUs, elastic straps, Minix and a TV remote controller.

The validation protocol consisted of several steps. At the beginning the sensor were turned on and left on the table to warm up for about ten minutes to reduce temperature effects on the sensor's readings [20]. After that it was possible to compute the gyroscope bias with 221e software that was removed to improve the estimation of angular velocity. Meanwhile the 49 markers were prepared, so each ball was attached on a piece of bio-adhesive tape that then was placed on the subject's skin. After that the SP system was prepared using Nexus software and following three fundamental steps before to start the acquisitions.

- Mask cameras was the first operation after creating the file related to the subject under analysis and it was necessary to mask any light reflections.

- Full calibration which allowed to the optoelectronic system to find the relationship between the coordinates of the camera image plane (2D) and the coordinates X, Y and Z of the marker in space (3D) and consisted in showing the active wand to all cameras to determine the geometric and optical characteristics of cameras (internal parameters) and the position and orientation of the camera frame relative to a certain laboratory frame (external parameters) [6], in this case the frame selected were 2500.
- Set origin for the reference system of laboratory. The wand was placed like showed in Figure 3.2

At this point the subject was prepared with marker to obtain the joint kinematic reference and with five IMUs; 1,2,5,6,7 respectively on foot, hand, wrist, upper arm and thorax (Figure 3.6) at the beginning for upper limb exercises and then with the same IMUs but for 2, on foot, ankle, thigh and pelvis (Figure 3.7) for lower limb exercises. Before starting the acquisitions, a calibration refinement was always performed.

Now was possible to acquire the subject firstly in a static position for a few seconds with left leg forward and left arm at shoulder height, as shown in Figure 3.5. This operation allowed to perform the manual labelling of marker set i.e., to associate the real marker to the light point returned by cameras. The asymmetry was needed to ease the identification of the right and left body sides. Then, a post-processing of the static recording consisting of three pipelines (KAD static processing, Plug in Gait Static and ULM_GPEM – Static pip) was executed on Nexus to calibrate the subject. It is important to make attention on the first pipeline cited that reports an acronym i.e., KAD (knee alignment device). The KAD in fact was a device with three marker which permitted to compute the rotation center of the knee [21] and in these models used it was simulated with the medial marker. In addition, these could be removed during dynamic acquisitions. However, at this point to the subject was asked to perform the desired movements for dynamic acquisitions, starting with shoulder, wrist elbow and trunk exercises. After the shifting of the IMUs on the inferior arts and recording a second static trial only with sensors it was possible to perform trials involving knee, hip, and ankle. At the end of recordings data collected during all trials were extracted from each IMUs and saved in .txt files and then in Nexus all trials were processed, to obtain model outputs and to extract variables of interest into two files .txt (Angles and Forces) after the anthropometrics parameters were entered on the software.

Post-processing steps in Nexus were:

- Launch shortcut ‘Reconstruct and label’

- Gap filling to fill holes caused by marker occlusion
- Launch pipeline: Plug in Gait Dynamic + ULM_GPEM – Dynamic pip
- Launch Export ASCII to exported desired information



Figure 3.5: Static position

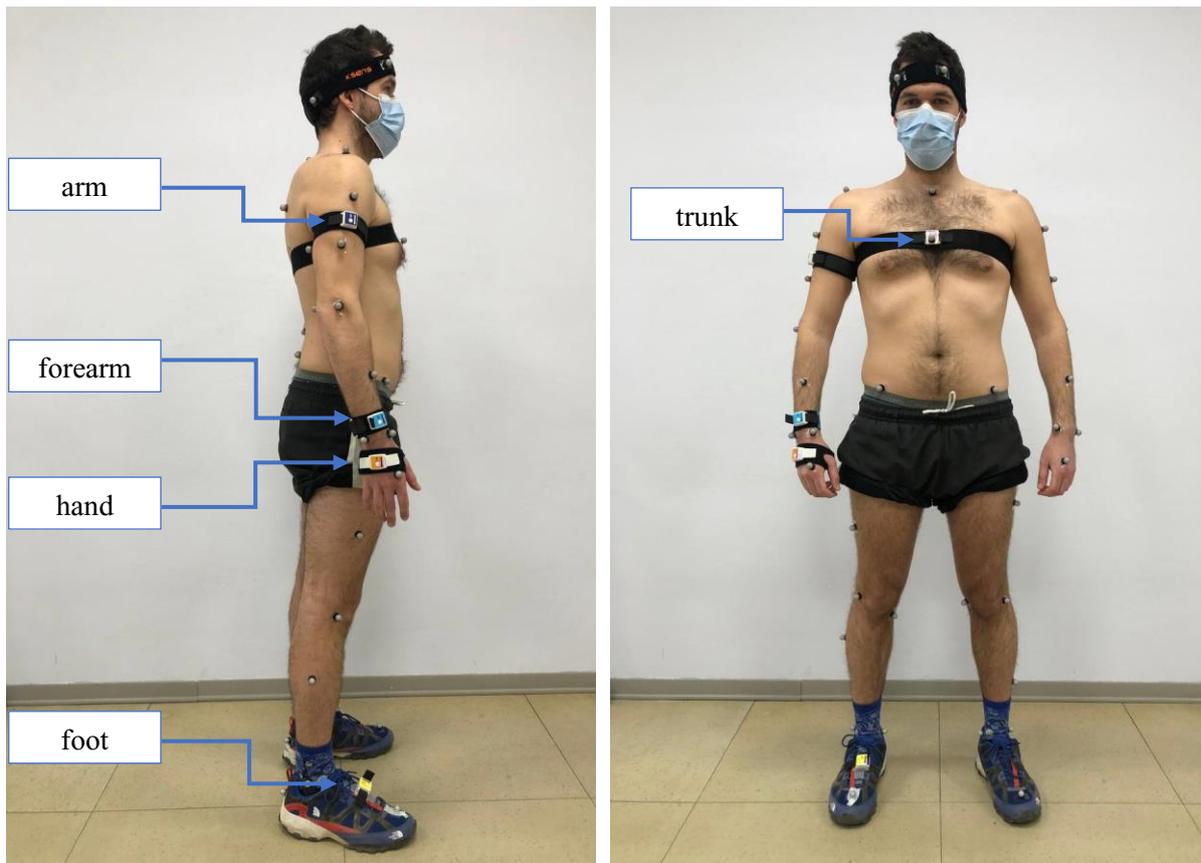


Figure 3.6 : The experimental equipment for upper limb exercises.

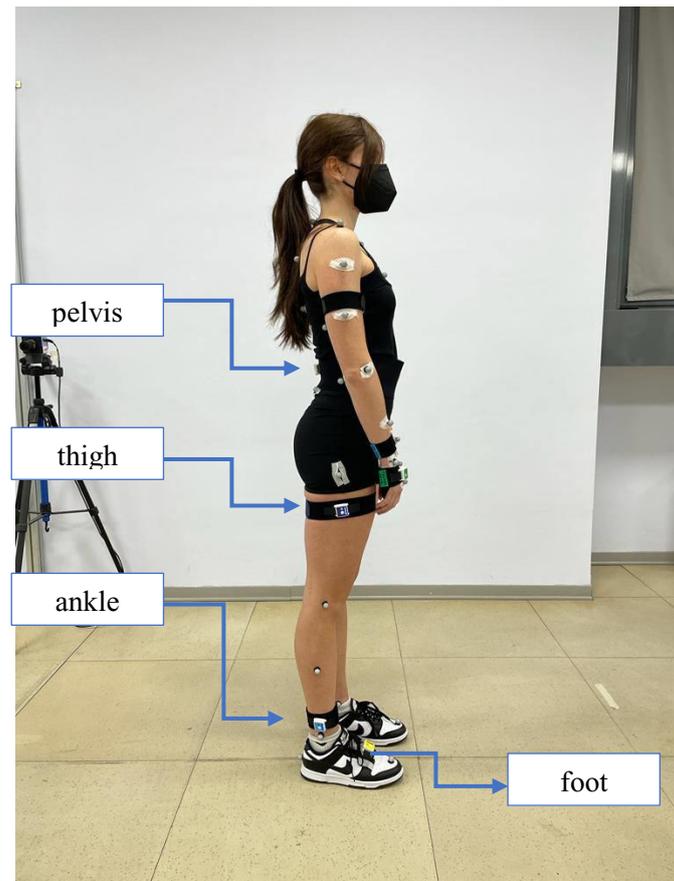


Figure 3.7 : The experimental equipment for trials involving lower limbs.

3.2 Experimental protocol

In this work nine healthy subject both females and males between 23 and 28 years (25 ± 2.1 y.o) were involved. They were called twice for a test and retest session and asked to perform a set of nine planar rehabilitation exercises that involved upper, lower limbs and the trunk. The exercises were the following, in order of execution (the first letter corresponds to the joint involved and the others refer to the movement performed):

1. SAA: shoulder Abduction/Adduction (AA)
2. WFE: wrist flexion/extension (FE)
3. EFE: elbow FE
4. TRV: Trunk rotation (R) around vertical axis (V)
5. TAP: Trunk R around antero-posterior axis (AP)
6. TFE: Trunk FE
7. KFE: Knee FE
8. HFE: Hip FE
9. AFE: Ankle FE

During first experimental session each subject performed twice three series of ten repetitions for each exercise. The three series were distinguished from each other by the amplitude of movement, these three amplitudes were:

- Small: $\sim 1/3$ ROM max
- Medium: $\sim 2/3$ ROM max
- Large: $\sim 3/3$ ROM max

Table 1: Overview of total recordings during test session for each subject; two is the number of recordings and three are the movement amplitudes requested.

	SAA	WFE	EFE	KFE	HFE	AFE	TRV	TAP	TFE	TOT
IMU	2x3	54								
SP	2x3	54								

One week apart was organized another recording session for the retest and the subject performed again twice the nine exercises listed before, but only one series of ten repetitions at $\sim 3/3$ of his maximum ROM.

Table 2 : Overview of total recordings during retest session for each subject; two is the number of recordings and three are the movement amplitudes requested.

	SAA	WFE	EFE	KFE	HFE	AFE	TRV	TAP	TFE	TOT
IMU	2x1	18								
SP	2x1	18								

The subjects were seated during each exercise, on a chair without a backrest and armrests but for the TFE where the subject was standing. He was in the centre of the capture volume and in front of the force plate that must be beat at the beginning and at the end of each trial. However, this type of chair permitted to occlude marker as little as possible.

The recordings were made with both systems IMU, and SP and the steps were:

1. Start acquisition with IMU software.
2. Start recording with Nexus.
3. Three beats on the force plate with the foot to identify the first synchronization point.
4. Performance of the first series of the first exercise (10 rep) keeping ROM about $1/3$ of the possible (small amplitude).
5. After about a pause of 10 seconds the subject repeated the same exercise, so 10 reps at medium amplitude trying to maintain ROM at about $2/3$ of his maximum.

6. Other 10 seconds of pause and then the subject performed 10 rep. at large amplitude, keeping the ROM as high as possible.
7. At the end, other three beats on the force pad with the foot in order to identify the second synchronization point, that will correspond with the end of the exercise.
8. Stop recording with Vicon.
9. Stop IMUs' acquisition.
10. All steps were repeated after starting a new recording and for each exercise.

The subjects started with the first six exercises of the list in the previous page, then the sensors were shifted on the lower limbs and the recording's steps (1-10) were repeated.

In the figure below an example of a subjects during execution of nine planar exercises and in particular for the exercise of the shoulder (on the left of Figure 3.8) and the large amplitude it was thought to place an obstacle to make the movement as repeatable as possible. Instead during wrist exercise was used a table to support the elbow because it was important that it did not elevate. For the retest session the steps were the same except for 5 and 6 that did not perform.

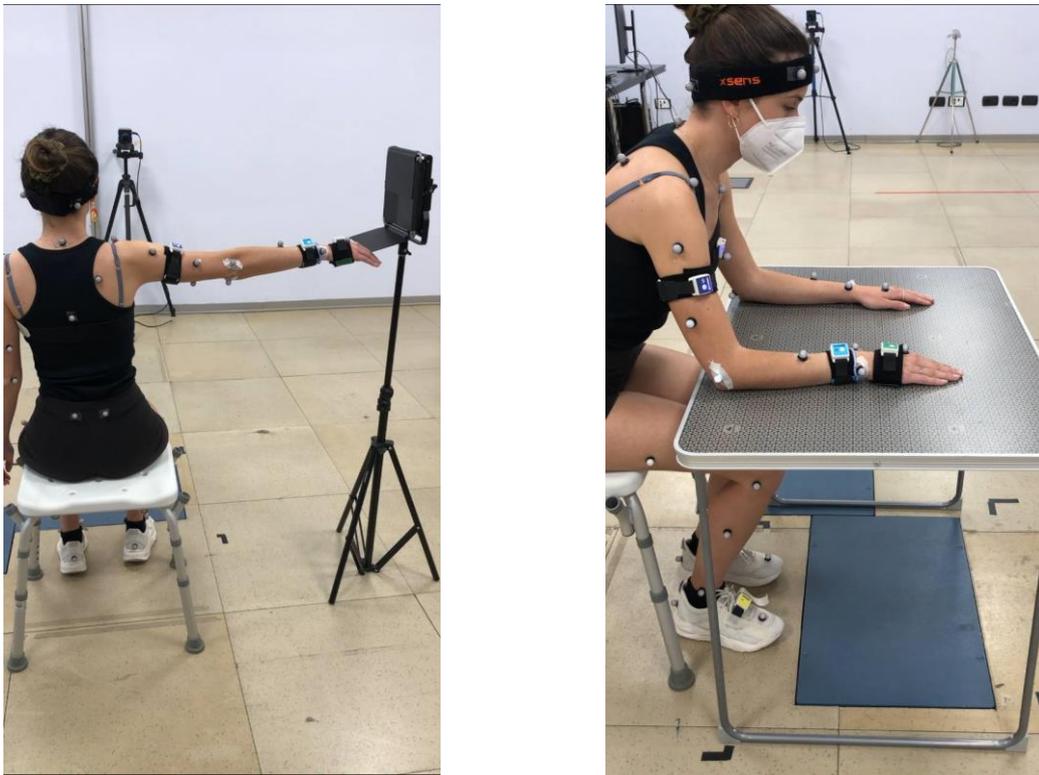


Figure 3.8 : Shoulder exercise (on the left) and wrist exercise (on the right).



Figure 3.9 : From left to right the elbow flex/extension and the trunk rotation around vertical axis.

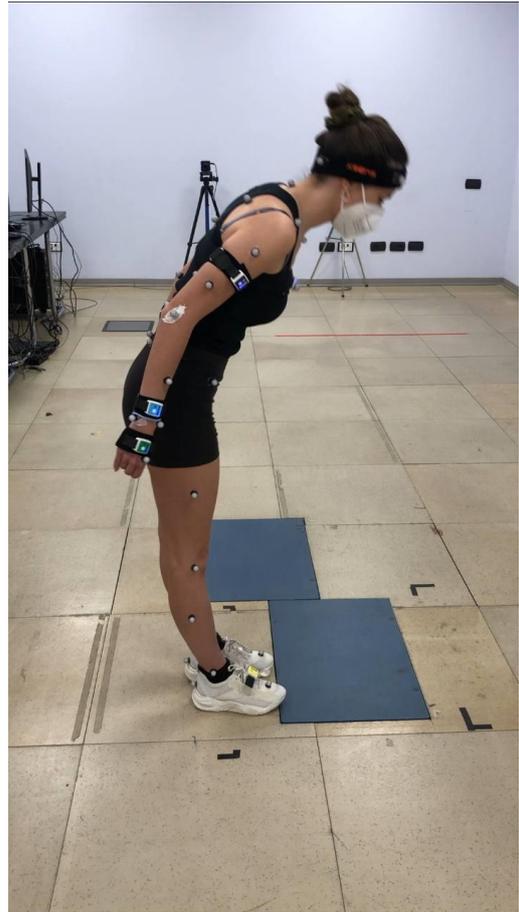


Figure 3.10: The trunk rotation around antero-posterior axis (on the left) and the trunk flex/extension (on the right).



Figure 3.11 : Respectively from left to right; knee, hip, and ankle flex/extension.

4 Data processing

4.1 Data pre-processing

The following processing was done on MATLAB (R2021b, The MathWorks Inc, Natick, MA, USA).

First, the synchronization between two systems is performed and so the peaks in the Figure 4.1 are used. The signals were cut between the third peak of the first group i.e., the green line that indicated the movement starting, and the first peak of the second group of peaks (red line i.e., the end of the movement).

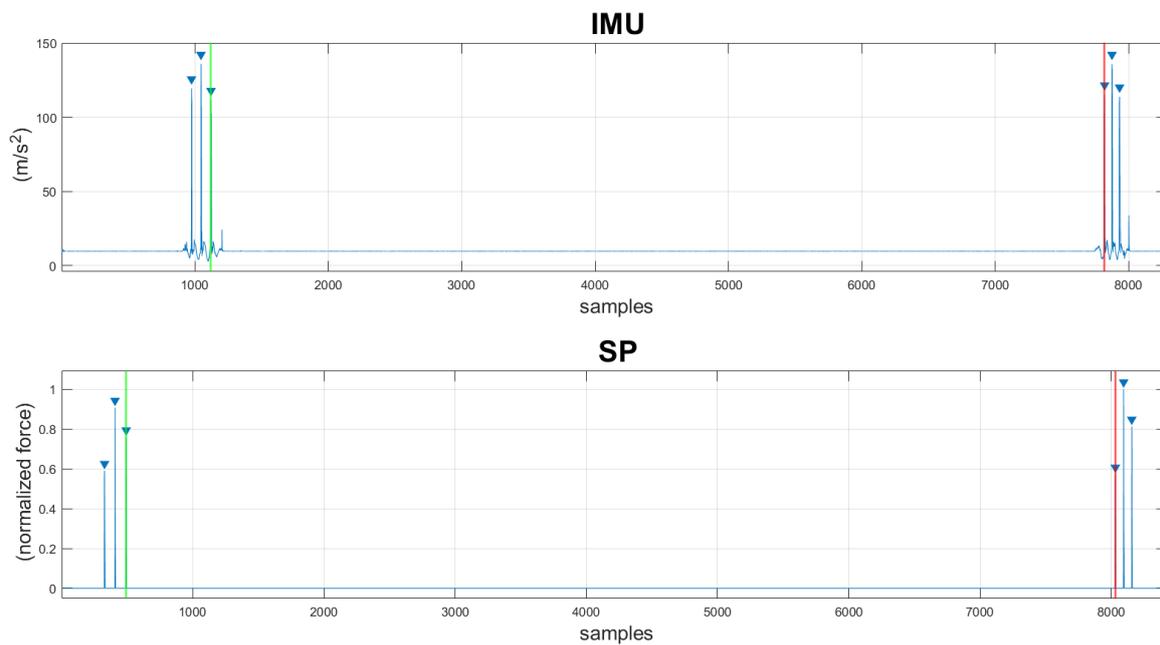


Figure 4.1: The peaks used to synchronize IMU (on the top) and SP.

IMU signals were cut, and resampled at 100 Hz, in this way the signal length was the same of the SP signal. Then data were segmented in three series of ten repetitions, filtering SP signals with a very low filter to obtain the envelope, which was made negative to make the segmentation easier.

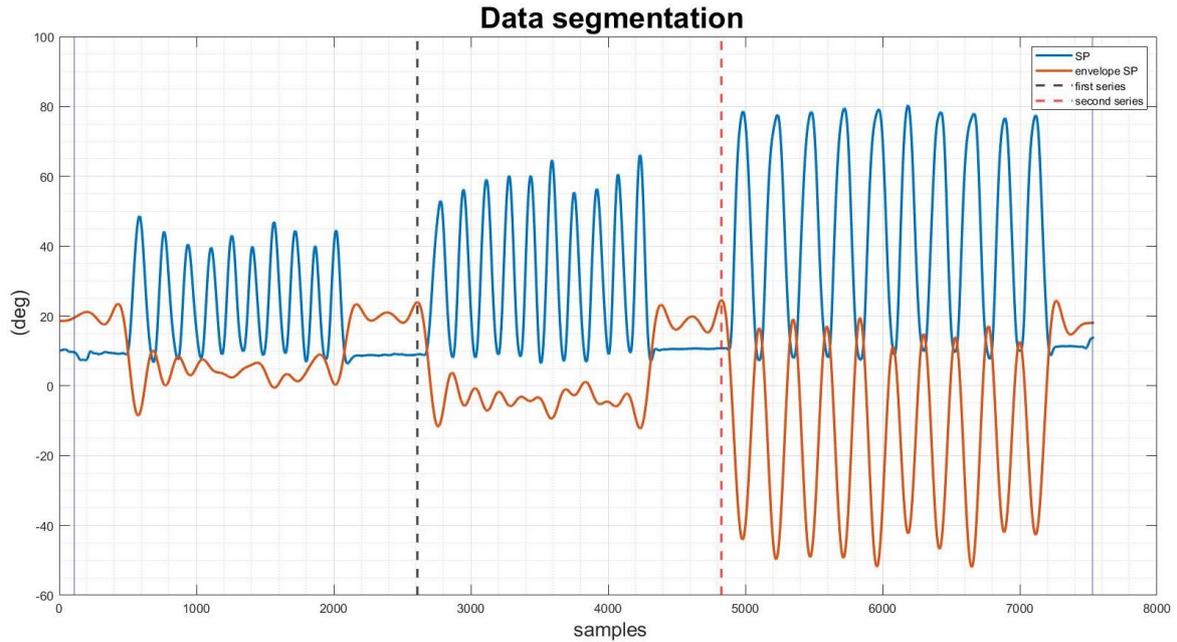


Figure 4.2: In blue SP signal and in orange SP envelope ($f_s=100$ Hz), the blue lines sign the start and the end of the movement, the black dotted line denotes the end of the first series and the red one signs the end of the second series.

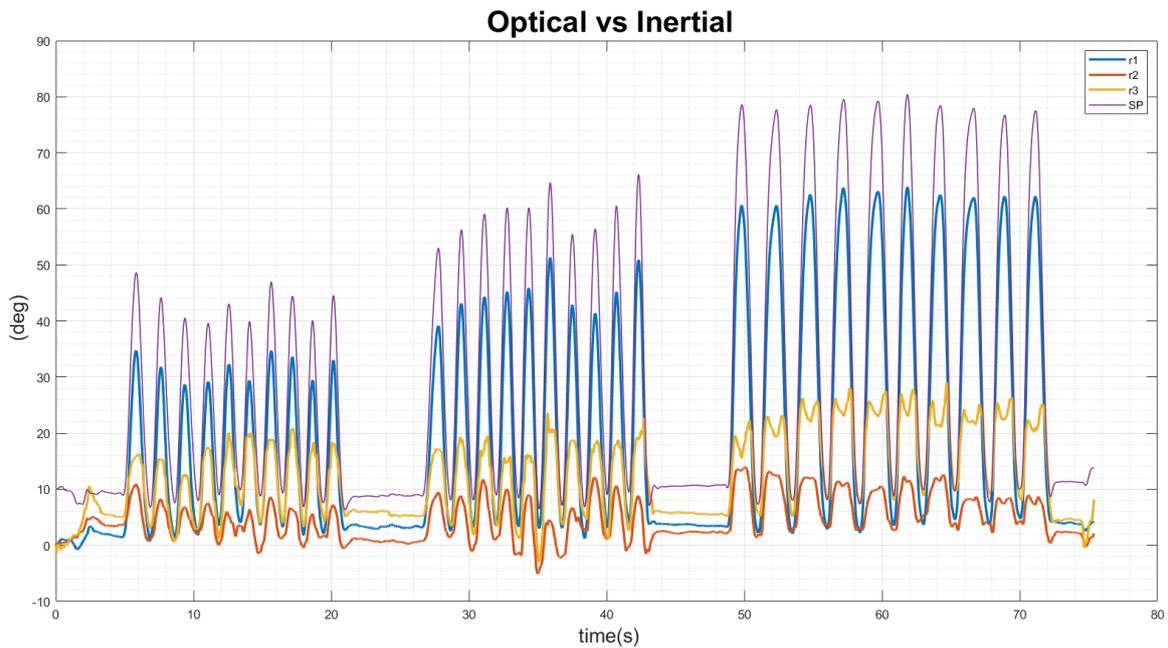


Figure 4.3: The plot shows SP and IMUs signals to check the synchronization.

To check the whole synchronization a plot with the IMU and SP signals was made. Now the synchronization had to be refined because the shot given by the foot was an impulse, so it had a larger bandwidth than the signal of interest, and the alignment was affected by uncertainty. To reduce the misalignment between the angle's series, the delay was changed,

and with this operation it was possible to minimize the objective function given by the RMS of the difference of the two angle's curves. This step which consisted in the translation on the time axis of some sample permitted to reduce error of some degrees like Figure 4.4 shows. The preparation of raw data was completed, and now it was possible to launch validation code, to count the repetitions and compute the quantities of interest for the comparison between the two systems and the statistics.

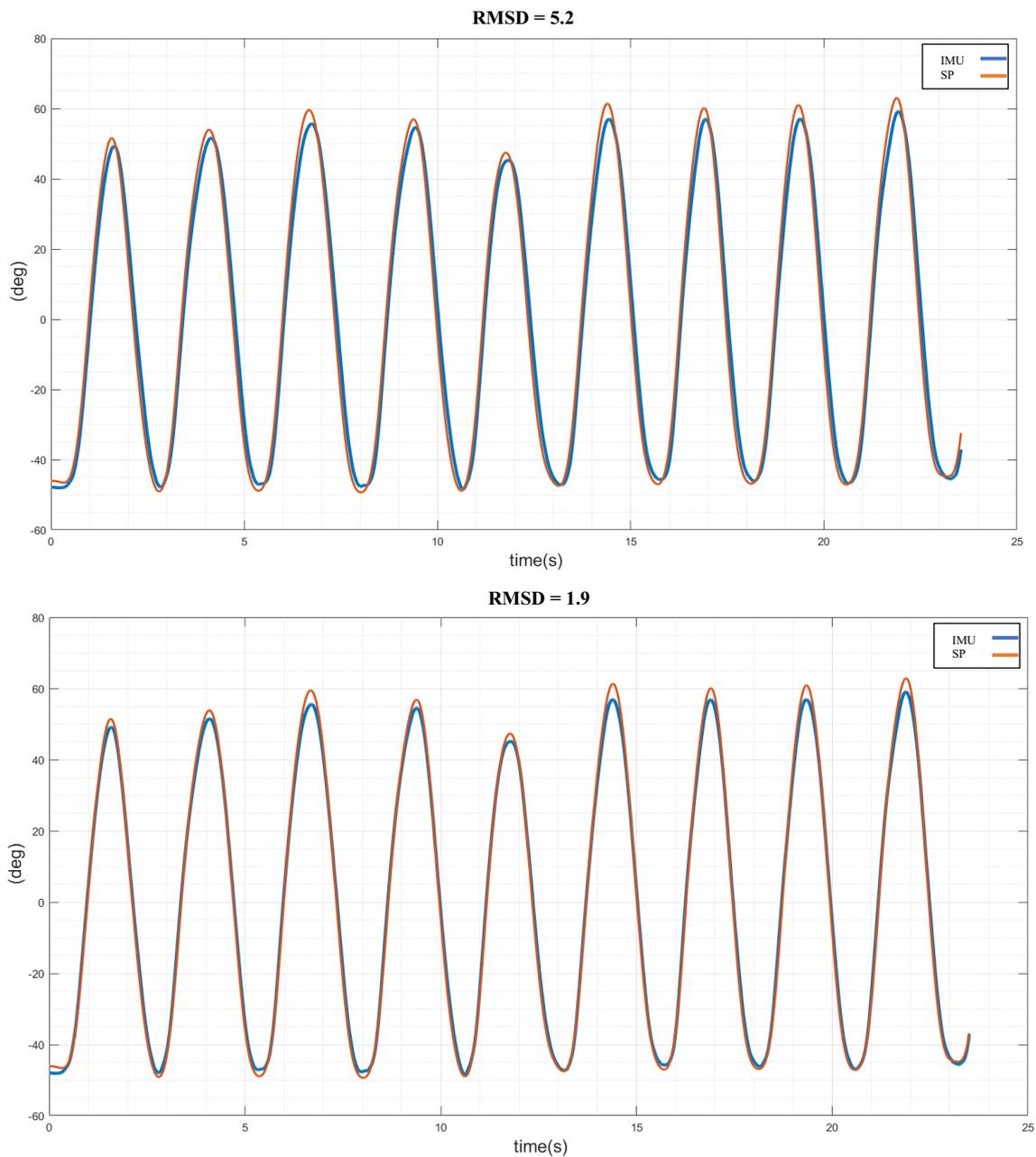


Figure 4.4: The misalignment reduction

4.2 Parameter extraction and metrics definition

4.2.1 Accuracy

The counting of the repetitions was made by the double threshold method, the beginning of each repetition was marked by the second threshold (th_2) crossing, the highest one, in the ascent time and the end of movement after the first threshold (th_1) crossing during the descent time, in other words the counting of one repetition occurred with the threshold crossing in the following order: th_1 , th_2 , th_2 , th_1 . As example a figure of a series of shoulder adduction/abduction is reported in Figure 4.5.

At the end with the angle time series extracted from both systems it was possible to calculate the RMSD after removal of offset, i.e., the mean value of the angle series. The offset was the result of the different definition of anatomical systems. The Figure 4.6 shows the difference and the RMSD value for a specific trial.

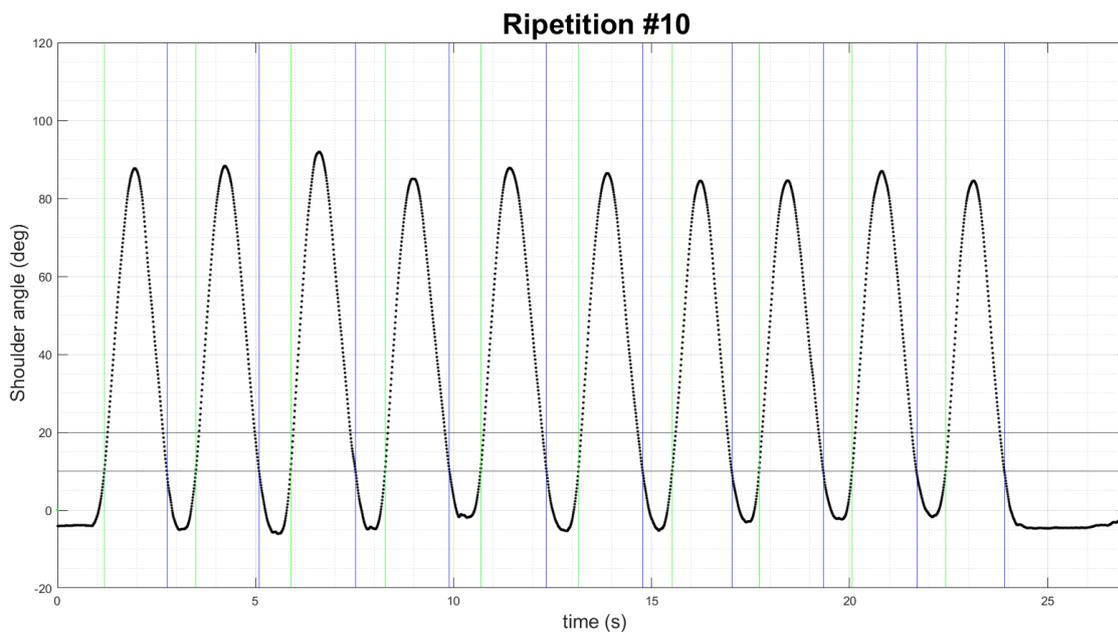


Figure 4.5: The figure is an example of validation code output: the count of the repetitions for shoulder exercise. The green line denotes the start of the repetition and the blue one the end of the same rep. The title updates when a new repetition is counted.

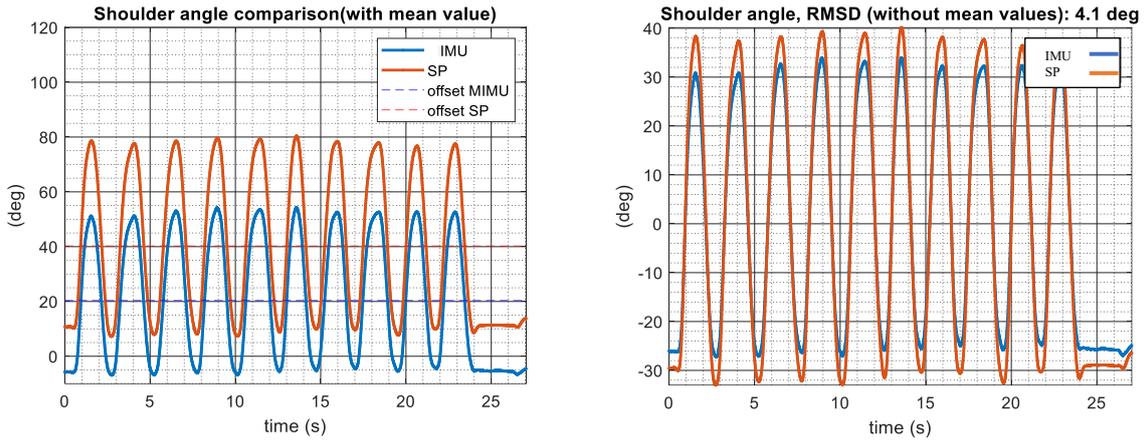


Figure 4.6: The angle waveforms with offset on the left and after the removal on the right.

The number of repetitions, the averaged ROM and execution time, and the RMSD were shown on the screen. The RMSD values of each exercise were saved for each subject. Keeping separate the three amplitudes the RMSD values of each subject and for each exercise were averaged to obtain nine mean values, corresponding to the nine planar exercises computed. In this way it was possible to have a clearer overview of mean errors between the two systems.

4.2.2 Reliability

The second aim of this thesis work was the computation of test-retest reliability to study the variability of the ROM due to the placement of the IMUs and SP markers on the subject's skin during the two experimental sessions. In fact, it is supposed that such variability can influence the ROM values. In particular, only the ROM related to the movement at the widest amplitude was considered for the test-retest analysis because it was the most repeatable. The test-retest method is the simplest method for testing the correlation between the results obtained from testing the same subjects twice in two different times. Theoretically, a test retaken after a month, for example, should give the same results of the first test performed one month before, to confirm that it was a reliable test. So, when a test-retest reliability was performed, some systematic differences in the population measurements between the first and second test could occur due to learning or habituation from repeated exposure to the measure's tests.

The computation of test-retest reliability was performed taking into account the twenty values obtained for each subject after the two recordings of each exercise performed at the same amplitude (one recording = 10 ROM values). These values were averaged to obtain a single mean value for each session (test and retest), namely $\mu_{i,j}$, where $i = \text{subject } \{\#1, \dots, \#9\}$, $j = \{1, 2\}$ for

test and retest, respectively. Table 3 shows the organization of the results obtained for test and retest sessions during each exercise.

Table 3: $\mu_{i,j}$ = the mean value of each matrix M , i =subject, j = experimental session (1=Test, 2= Retest).

	Test session		Retest session	
N subjects	ROM [1x20]	μ_1 [1x1]	ROM [1x20]	μ_2 [1x1]
#1	$M_{1,1}$	$\mu_{1,1}$	$M_{1,2}$	$\mu_{1,2}$
#2	$M_{2,1}$	$\mu_{2,1}$	$M_{2,2}$	$\mu_{2,2}$
#3	$M_{3,1}$	$\mu_{3,1}$	$M_{3,2}$	$\mu_{3,2}$
#4	$M_{4,1}$	$\mu_{4,1}$	$M_{4,2}$	$\mu_{4,2}$
#5	$M_{5,1}$	$\mu_{5,1}$	$M_{5,2}$	$\mu_{5,2}$
#6	$M_{6,1}$	$\mu_{6,1}$	$M_{6,2}$	$\mu_{6,2}$
#7	$M_{7,1}$	$\mu_{7,1}$	$M_{7,2}$	$\mu_{7,2}$
#8	$M_{8,1}$	$\mu_{8,1}$	$M_{8,2}$	$\mu_{8,2}$
#9	$M_{9,1}$	$\mu_{9,1}$	$M_{9,2}$	$\mu_{9,2}$

Then, to assess the test-retest reliability, μ_1 and μ_2 were organized in a matrix (Table 4) and starting from the $\mu_{i,j}$ measurements it is possible to compute the terms required by the intraclass correlation coefficient (ICC) formulation shown in (4.2.1) proposed by McGraw, 1996 [22].

Table 4: Data matrix to compute ICC.

N subjects	μ_1	μ_2
#1	$\mu_{1,1}$	$\mu_{1,2}$
#2	$\mu_{2,1}$	$\mu_{2,2}$
#3	$\mu_{3,1}$	$\mu_{3,2}$
#4	$\mu_{4,1}$	$\mu_{4,2}$
#5	$\mu_{5,1}$	$\mu_{5,2}$
#6	$\mu_{6,1}$	$\mu_{6,2}$
#7	$\mu_{7,1}$	$\mu_{7,2}$
#8	$\mu_{8,1}$	$\mu_{8,2}$
#9	$\mu_{9,1}$	$\mu_{9,2}$

As anticipated, the ICC is the most common coefficient to assess the reliability. However, the ICC is characterized by many definitions. Currently, two main nomenclatures can be found in literature: Shrout and Fleiss (1979) proposed six ICC definitions, while McGraw and Wong (1996) proposed ten ICC forms. All these conventions are based on the mean square values derived from the analysis of variance (ANOVA). Table 5, taken from Terry K. Koo, 2016 [23], reports both forms of ICC to better understand the differences and choose the correct ICC to be used as a reference.

Table 5: Equivalent ICC forms between Shrout and Fleiss (1979) and McGraw and Wong (1996).

McGraw and Wong (1996)	Shrout and Fleiss (1979)	Formulas for calculating ICC
One-way random effects, A, 1	ICC (1,1)	$\frac{MSR - MSW}{MSR + (k + 1)MSW}$
Two-way random effects, C, 1	-	$\frac{MSR - MSE}{MSR + (k - 1)MSE}$
Two-way random effects, A, 1	ICC (2, 1)	$\frac{MSR - MSE}{MSR + (k - 1)MSE + \frac{k}{n}(MSC - MSE)}$
Two-way mixed effects, C, 1	ICC (3, 1)	$\frac{MSR - MSE}{MSR + (k - 1)MSE}$
Two-way mixed effects, A, 1	-	$\frac{MSR - MSE}{MSR + (k - 1)MSE + \frac{k}{n}(MSC - MSE)}$
One-way random effects, A, k	ICC (1, k)	$\frac{MSR - MSW}{MSR}$
Two-way random effects, C, k	-	$\frac{MSR - MSE}{MSR}$
Two-way random effects, A, k	ICC (2, k)	$\frac{MSR - MSE}{MSR + \frac{MSC - MSE}{n}}$
Two-way mixed effects, C, k	ICC (3, k)	$\frac{MSR - MSE}{MSR}$
Two-way mixed effects, A, k	-	$\frac{MSR - MSE}{MSR + \frac{MSC - MSE}{n}}$

*A = absolute agreement, C= consistency, 1= single rater/measurement, k= multiple raters/measurements
1 = one-way random effects model, 2= two-way random effects model, 3= two-way mixed effects model*

Based on the ANOVA repeated measures definitions, the MSR is the mean square for rows, the MSC is a mean square for columns, and MSE is a residual mean square traditionally referred to as the mean square of errors, n is the number of observations, in this case $n = N$ subjects [22]. A more detailed explanation of the ANOVA and its link with the ICC can be found in the Appendix. More in detail, regarding the two ICC nomenclatures reported in Table 5, McGraw and Wong based the choice of the suitable computational formula among the ten forms of ICC on the “model”, the “type” and the “definition”:

- The model can be selected among 1-way random effects, 2-way random effects and 2-way mixed effects.
- The “type” can be selected between single rater/measurement and the mean of k rater/measurements. The rater is the person who conducts the test.
- The “definition” can be selected between consistency and absolute agreement.

Shrout and Fleiss presented six forms with two numbers in parentheses, the first number is the model and the second one is the type. Each choice is correlated with the planned experimental design and to the type of reliability to be assessed, which can be inter-rater reliability, intra-rater reliability or test-retest reliability as in this case. The reference work of *Terry K. Koo, 2016* reports the suggestion of Shrout and Fleiss to select 2-way mixed effects model in test-retest reliability study because the repeated measurements are not randomized samples. Furthermore, for test-retest study, absolute agreement definition should always be used because measurements are meaningless if there is no agreement between repeated measurements performed. Another help that could guide the researchers to the appropriate selection of ICC is the flowchart shown in Figure 4.7 adapted from, *Terry K. Koo, 2016*.

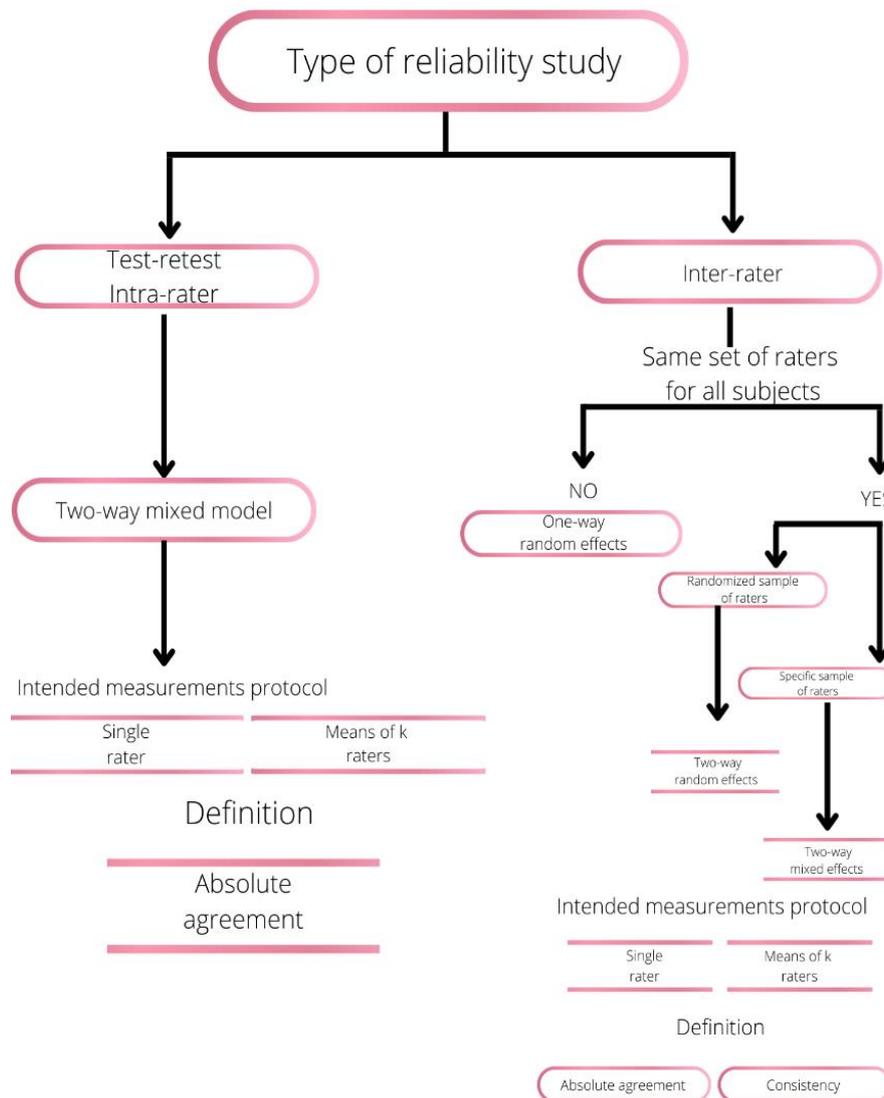


Figure 4.7: Flowchart to guide researchers in an appropriate selection of ICC (adapted from Terry K. Koo, 2016).

The flowchart starts with the type of study conducted and its levels lead to the decision of the model, of the type and then of definition. It is important making the right choice of the ICC to be applied on the specific study case, but it is not simple since most of the literature deals with rater reliability and not with test-retest reliability of physical performance measure [24].

After having decided the type of ICC to be used in the specific study design it must be properly interpreted. Generally, the ICC value is bounded between 0 and 1, an ICC equal to 0 indicates absence of reliability while an ICC equal to 1 indicates a perfect reliability. In particular, *Terry K. Koo, 2016* proposed the following intervals to associate each ICC value with a level of reliability:

Table 6: ICC intervals by Koo et al., 2016.

ICC	Reliability
< 0.5	poor
≥0.5 & <0.75	moderate
≥0.75 & <0.90	good
≥0.90	excellent

Since the aim of this thesis was the assessment of test-retest reliability the “two-way mixed effects” model was chosen in this study, following the scheme in Figure 4.7. This model was preferred over the one-way model because time was a fundamental design factor in the proposed test-retest assessment (i.e., the two time points were not interchangeable). Moreover, the chronology was important to detect some differences among the outcome measures. In fact, in this case an ICC computed using the one-way model would underestimate the reliability. In addition, a mixed-effect model was recommended over a random effect model and the reason was in the test and retest time points that were specified and identical across all study subjects [25]. Furthermore, the type based on mean of k-measurements was the most suitable for this study, since it was performed the mean of 20 values of the ROM for each experimental session. At the end, the flowchart, following this experimental design, recommends the absolute agreement definition. All these choices resulted in the ICC(A, k) (i.e., the ICC(3, k) according to Shrout and Fleiss) which was the most appropriate to pursue the objective of this thesis. So, the final computational formula used was:

$$ICC(3, k) \equiv ICC(A, k) = \frac{MSR - MSE}{MSR + \frac{MSC - MSE}{n}} \quad (4.2.1)$$

In this thesis, the ICC was computed in MATLAB R2021b with a specific function [26], and it was obtained for each exercise. This allowed to take into account variability regarding the anatomical calibration procedure, the presence of soft tissues artefacts, and errors due to the IMU orientation reconstruction. The function requested as input the data matrix shown in Table 4, the type of the ICC to be computed, and the level of significance ($\alpha = 0.05$).

4.2.2.1 Test-retest reliability – a literature review

Test-retest reliability of an instrument is computed by measuring subjects at two distinct occasions (T1 and T2), under the same conditions with that specific instrument under analysis and then computing the correlation. It evaluates the degree of agreement of each participant's scores on both times. It has been used in similar research and is highly recommended when the goal of the study is the assessment of reliability using one evaluator. Then, if the correlation is large, is obvious that the test-retest is good.

The test-retest analysis is routinely recommended during the validation phase of many measurement instrumentations, like the IMU-based system (DoMoMEA) in this thesis work.

In literature a lot of studies use test-retest to assess an instrument, and to test the agreement between two measures obtained in two or more different sessions of acquisition. Several studies used the ICC and involved a great number of subjects to obtain reliable results. The *S. Schneiberg et al., 2010* [27] study for example performed the test-retest reliability of kinematics measures in some children affected by cerebral palsy. The main aim of the study was to evaluate the ability of the subjects involved to grasp an object situated at three different distances (d1, d2, d3). The parameters were extracted from thirteen children of about 9 years old. They were evaluated three times over 5 weeks by the same observer. The exercises were asked to perform were sagittal trunk displacement, elbow extension, shoulder flexion and shoulder abduction/adduction. The study determined the test-retest reliability implementing an ICC model (2, k), and the values obtained suggested for all kinematic variables and all three targets a moderate/excellent reliability except for shoulder horizontal abduction/adduction for d1 and shoulder flexion for d2. Since the purpose of this study was to find kinematic variables that might be utilized as reliable outcome measures in randomized clinical trials of upper limb therapies in children with cerebral palsy, so the ICC values were acceptable when exceed 0.80 and among the variables rated the trunk displacement, and the elbow extension were given good

results so these two parameters will be reliable indicators of change in movement quality for this or a similar upper limb intervention. But for the shoulder limits a reason could be the control of shoulder initial position that was difficult to repeat. Overall, the main limitation of the study are the kinematic variables which are task specific so the reliability should be interpreted in the context of task requirements. *D. Laroche et al., 2011* [28] investigated the test-retest reliability of 3D gait analysis (3DGA) in hip osteoarthritis patients and involved 23 patients in two sessions of gait analysis. The aims were three, but we took more attention on the assessment of reliability. In this work was suggested the use of two-way intraclass correlation coefficient to assess the reliability between sessions. The results demonstrated that all variables and most of kinematic joint angles presented good to excellent reliability with ICC values between 0.7 and 0.9 except for pelvic angles and knee frontal movement, which did not demonstrate acceptable reliability because their ICCs were lower than 0.7. For the hip this lower ICCs could find an explanation in the difficulty to identify the anatomical landmarks around the hip and pelvis and this could take some errors in anthropometric measurements. For the knee it could be explained by the small ROM of the knee in the frontal plane. The study cited had a main limitation, that was the small number of patients involved in the study and the asymmetric male to female distribution. This is a common limit among several studies conducted in this research field.

Another interesting study was that of *Nilsson et al., 2022* in which 22 healthy subjects were analysed during two sessions of gait analysis, including comfortable, fast gait and stair walking, dressed with 12 wearable inertial sensors in feet, shank, thigh, pelvis, thorax, and arms. The aim was not only test gait speed, but it was extended to extraction of kinematic and temporal measures from the lower body and arms, based on data from a wearable inertial sensor system. The ICC (3, k, absolute agreement) was the metrics used to quantify how much the extracted parameters i.e., kinematic, and temporal outcome measures were reliable. Overall, the ROM showed the best reliability, maybe for the assumption that all joints are aligned at zero angles in the standardized position. But the suggestion was to include in the further studies pathologic patients with some disorders to test the effective reliability of the ROM in all joints. In literature there are few studies which focused on upper limb motion analysis, but it is very important in neurological and musculoskeletal disorders, e.g., stroke or Parkinson's disease and fundamental for clinicians to make a decision about a rehabilitation protocol or intervention.

Generally, there are many factors that can influence reliability such as the conditions between test and retest, or the test procedures applied by the evaluator or also factors like learning

fatigue and motivation effects [29]. In the present study the test procedures were standardized as much as possible.

Cai et al., [30] instead computed test-retest reliability of upper limb kinematics, involving ten healthy male subjects performed a battery of upper limb movements and measured with Microsoft Kinect v2 and the Vicon motion capture system. These participants were called for two testing sessions one week apart and after the performing of four exercises the ROM was extracted and used to comparison two systems and to assess the reliability of the most recent one, the Microsoft Kinect. The results demonstrated a good reliability of the ROM because ICC(3,k) values were between 0.68 and 0.96, but it was noted that the reliability could be task-dependent and plane-dependent, in fact the repeatability of upper limb motions in the frontal and transverse planes was lower than that in the sagittal plane. In addition, there were some differences of the results between sessions maybe due to the choice of starting and ending point of the movement that should be standardized to make the movement more repeatable.

At the end of this brief overview of other studies which performed the statistical analysis based on test-retest the main observation is that it is much dependent to the measurement instrument with which the parameters under study, such as ROM were extracted and that there are some limitations for exercises involving shoulder or hip. The studies present different models of ICC used for the assessment of the reliability, but the most frequent are ICC(2,1), ICC(3,1) and ICC(3,k).

5 Results

In this section were reported the results obtained from this study. The accuracy of joint kinematics like explained was expressed by the means of the RMSD. In the tables were shown the RMSD computed in each recording for each subject. It was preferred to report only values obtained during test and involving retest only to assess the reliability, because during test were performed much more series of the exercises (1/3, 2/3, 3/3 ROM max) and the analysis resulted more complete. Then, for each exercise the values were averaged, and the result was reported in the last row with a respective standard deviation.

Table 7: Accuracy evaluation; RMSD values for shoulder, wrist, and elbow. S1= first recording at 1/3 ROM max S2= second recording at 1/3 ROM max.

SUBJ	SAA		WFE		EFE	
	S1	S2	S1	S2	S1	S2
1	1.7	3.4	0.8	0.7	11	15
2	2.5	2.7	1.3	3.4	5.4	8.6
3	3.7	2.9	0.6	0.6	5.6	7.1
4	2.5	2.8	2.1	2.1	6.0	6.6
5	1.1	0.8	0.7	0.1	11	3.2
6	1.6	3.0	1.1	2.2	7.7	7.0
7	3.3	4.1	1.9	1.1	6.5	6.4
8	2.6	3.1	1.7	0	2.6	4.1
9	1.9	1.9	1.4	1.2	7.1	7.1
MEAN	2.5 ± 0.9		1.3 ± 0.8		7.1 ± 2.9	

Table 8: Accuracy evaluation; RMSD values for knee, hip, and ankle. S1= first recording at 1/3 ROM max S2= second recording at 1/3 ROM max.

SUBJ	KFE		HFE		AFE	
	S1	S2	S1	S2	S1	S2
1	2.9	2.8	1.1	0.8	3.5	3.5
2	3.2	4.1	4.2	3.1	2.8	1.7
3	1.6	1.4	2.1	2.4	1.4	3.9
4	2.3	3.4	5.2	5.3	1.1	2.9
5	0.9	1.2	1.5	1.4	1.8	1.9
6	1.3	1.4	2.9	3.0	0.9	1.4
7	5.1	5.7	1.8	1.9	0.9	0.7
8	2.7	3.0	4.2	3.7	0.8	0.5
9	3.3	2.9	3.8	3.9	1.9	2.1
MEAN	2.7 ± 1.3		2.9 ± 1.3		1.9 ± 1.0	

Table 9: Accuracy evaluation; RMSD values for trunk rotation around vertical and antero-posterior axes and trunk flex/extension. S1= first recording at 1/3 ROM max S2= second recording at 1/3 ROM max.

SUBJ	TRV		TAP		TFE	
	S1	S2	S1	S2	S1	S2
1	2.8	0.4	0.6	0.7	1.0	0.7
2	2.6	2.8	1.2	0.7	1.0	1.3
3	3.8	0	2.0	1.5	1.6	0.8
4	1.4	1.2	1.0	1.0	0.9	1.0
5	1.0	1.3	1.1	0.9	0.9	1.6
6	1.5	1.9	1.0	1.1	0.6	1.1
7	2.2	3.7	1.3	2.1	1.3	2.2
8	0.2	0.6	1.1	0.8	1.0	0.7
9	2.3	2.4	0.8	0.7	1.5	1.2
MEAN	1.8 ± 1.1		1.1 ± 0.4		1.1 ± 0.4	

Table 10: Accuracy evaluation; RMSD values for shoulder, wrist, and elbow. M1= first recording at 2/3 ROM max M2= second recording at 2/3 ROM max.

SUBJ	SAA		WFE		EFE	
	M1	M2	M1	M2	M1	M2
1	2.4	3.9	0.5	0.8	11.0	15.0
2	3.7	4.0	2.0	3.8	8.9	9.4
3	4.5	5.7	1.0	0.8	7.9	9.3
4	2.8	3.1	3.1	3.0	5.6	7.0
5	2.7	2.0	1.5	1.0	4.2	10.8
6	4.1	4.4	1.8	0.8	14.2	9.3
7	4.4	5.3	4.1	2.1	11.7	16.9
8	3.8	3.4	2.1	2.4	4.7	5.7
9	2.9	3.3	2.3	2.9	8.3	7.6
MEAN	3.7 ± 0.9		2.0 ± 1.1		9.3 ± 3.4	

Table 11: Accuracy evaluation; RMSD values for knee, hip, and ankle. M1= first recording at 2/3 ROM max M2= second recording at 2/3 ROM max.

SUBJ	KFE		HFE		AFE	
	M1	M2	M1	M2	M1	M2
1	3.7	3.4	1.1	1.2	2.8	2.8
2	3.1	5.5	5.0	4.6	3.0	2.4
3	2.7	1.7	3.9	3.6	1.2	0.8
4	2.3	2.3	6.0	7.0	1.1	2.3
5	1.3	1.5	3.5	2.4	1.3	1.6
6	2.0	2.0	3.7	3.7	0.9	0.9
7	7.3	7.3	2.7	3.1	1.0	1.2
8	3.7	3.7	2.9	2.6	0.7	0.5
9	3.9	3.3	4.6	4.9	2.2	2.3
MEAN	3.4 ± 1.7		3.7 ± 1.5		1.6 ± 0.8	

Table 12: Accuracy evaluation; RMSD values for trunk rotation around vertical and antero-posterior axes and trunk flex/extension. M1= first recording at 2/3 ROM max M2= second recording at 2/3 ROM max.

SUBJ	TRV		TAP		TFE	
	M1	M2	M1	M2	M1	M2
1	3.5	4.1	0.7	0.9	1.3	0.9
2	2.4	3.1	2.3	0.8	1.5	1.8
3	4.2	4.0	1.8	2.2	1.9	0.9
4	1.8	1.8	1.5	1.2	1.6	1.3
5	1.6	2.1	2.0	1.5	1.9	1.4
6	2.1	2.4	1.6	1.4	0.8	0.7
7	3.3	4.6	2.1	2.9	1.6	3.1
8	1.4	0.8	1.6	1.2	0.9	0.7
9	2.5	2.4	0.8	1.1	1.6	1.2
MEAN	2.7 ± 1.1		1.5 ± 0.6		1.4 ± 0.6	

Table 13: Accuracy evaluation; RMSD values for shoulder, wrist, and elbow. L1= first recording at 3/3 ROM max L2= second recording at 3/3 ROM max.

SUBJ	SAA		WFE		EFE	
	L1	L2	L1	L2	L1	L2
1	5.5	5.2	0.9	1.1	11.2	12.2
2	5.6	4.8	3.7	4.1	23.0	22.0
3	7.4	7.8	2.5	2.4	10.4	12.8
4	3.6	4.1	3.6	3.1	5.7	6.0
5	6.8	5.7	3.1	3.6	11.0	13.0
6	8.5	8.9	3.4	6.3	22.5	13.0
7	4.9	5.1	8.9	3.8	11.1	21.8
8	5.2	4.5	4.6	5.4	7.3	5.6
9	3.9	4.2	3.4	2.5	10.8	10.9
MEAN	5.7 ± 1.5		3.7 ± 1.8		12.8 ± 5.6	

Table 14: Accuracy evaluation; RMSD values for knee, hip and ankle. L1= first recording at 3/3 ROM max L2= second recording at 3/3 ROM max.

SUBJ	KFE		HFE		AFE	
	L1	L2	L1	L2	L1	L2
1	4.4	4.2	1.7	1.8	3.4	3.3
2	4.4	6.3	6.6	5.5	3.7	3.8
3	2.2	1.8	3.6	3.6	1.3	1.4
4	5.0	4.2	7.3	8.5	1.7	2.6
5	2.1	1.9	4.0	4.3	2.5	1.8
6	2.9	2.9	4.5	4.4	0.7	0.8
7	8.9	7.9	3.4	3.8	1.5	1.6
8	3.7	4.8	3.3	3.7	0.9	0.9
9	5.1	4.1	5.1	5.5	2.9	2.9
MEAN	4.3 ± 1.9		4.5 ± 1.7		2.1 ± 1.0	

Table 15: Accuracy evaluation; RMSD values for trunk rotation around vertical and antero-posterior axes and trunk flex/extension. L1= first recording at 3/3 ROM max L2= second recording at 3/3 ROM max.

SUBJ	TRV		TAP		TFE	
	L1	L2	L1	L2	L1	L2
1	4.6	4.5	1.3	1.5	2.0	1.5
2	3.6	3.6	3.1	1.3	2.6	2.8
3	5.1	5.2	2.9	2.9	2.5	1.4
4	0.1	2.3	2.0	1.7	2.2	2.6
5	2.8	3.2	0.7	0.9	1.4	1.8
6	3.0	3.5	2.6	1.7	1.6	1.1
7	4.5	6.6	2.7	3.9	1.8	3.9
8	1.4	0.9	1.7	1.2	0.9	1.1
9	3.0	2.8	1.2	1.4	1.9	1.5
MEAN	3.4 ± 1.6		1.9 ± 0.9		1.9 ± 0.7	

Now for a more representative and faster interpretation of the results six boxplots were represented, two for each amplitude of execution. In a boxplot were grouped the distributions of SAA, WFE, EFE, KFE, HFE, AKE. In the other boxplot instead were reported the distributions of TRV, TAP, TFE.

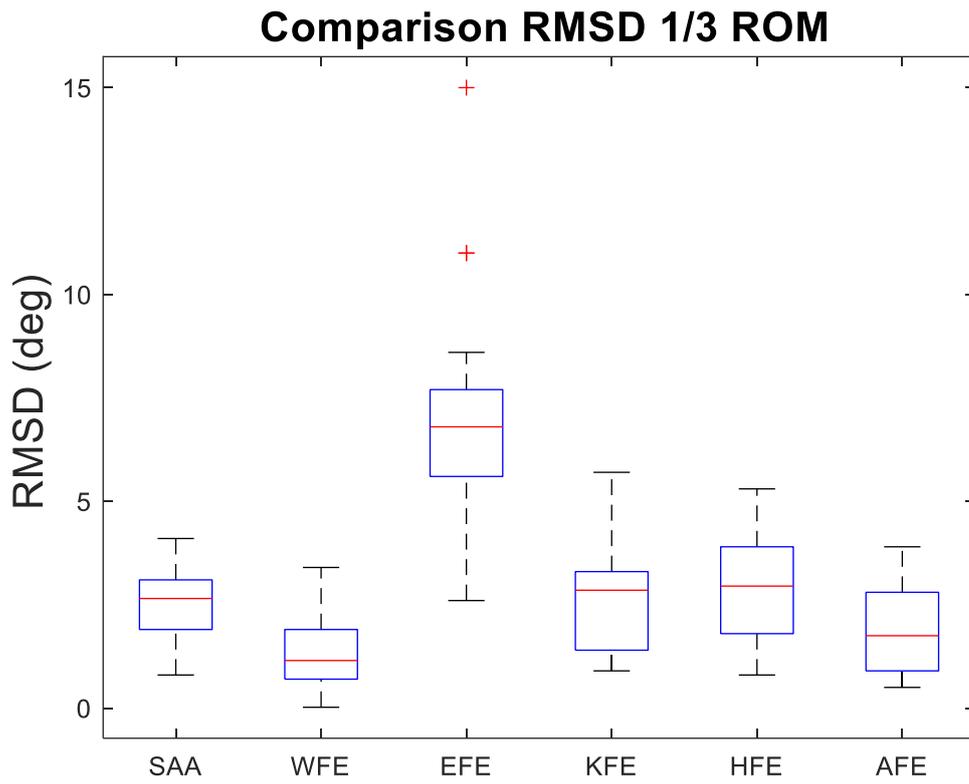


Figure 5.1: Boxplot of exercises involving upper and lower limbs - small amplitude.

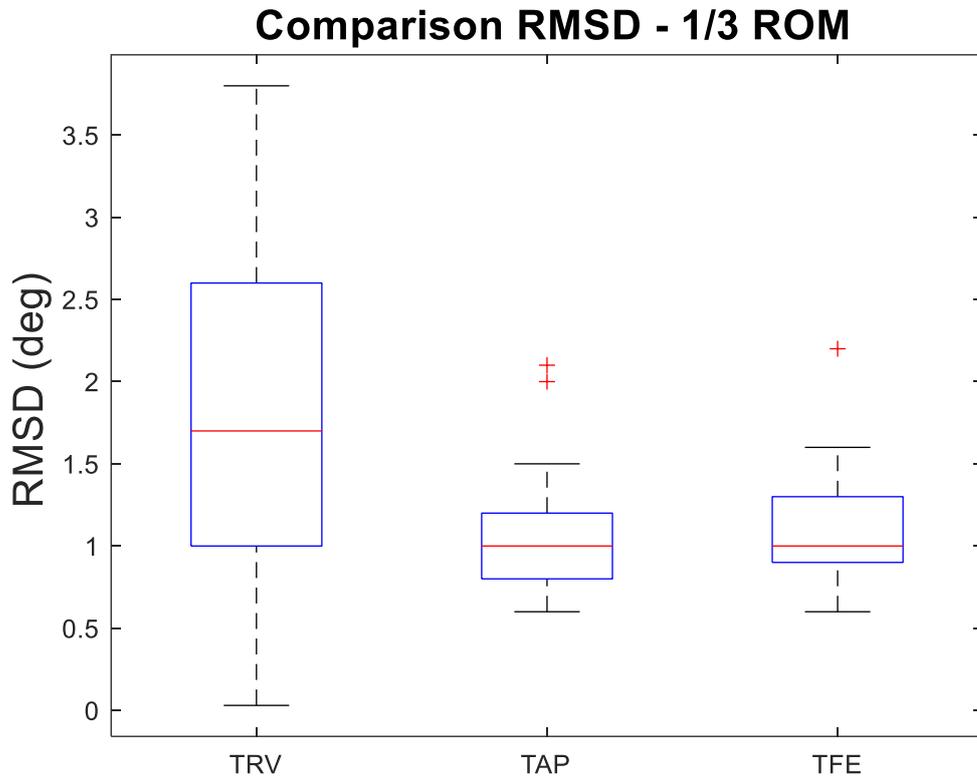


Figure 5.2: Boxplot of exercises involving trunk - small amplitude.

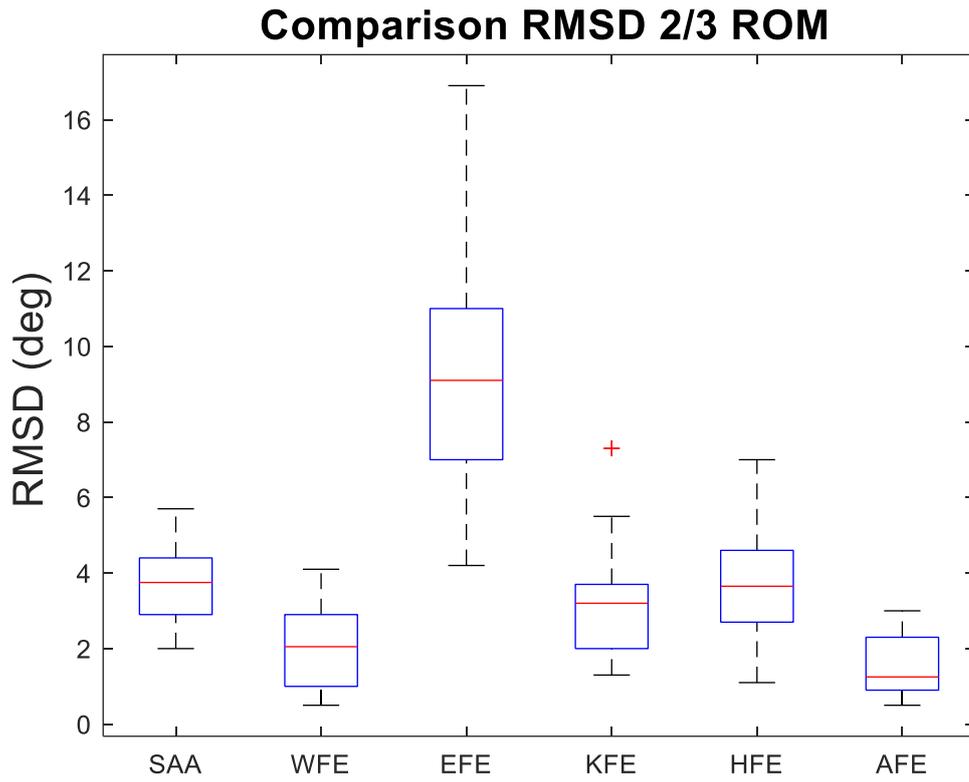


Figure 5.3: Boxplot of exercises involving upper and lower limbs - medium amplitude.

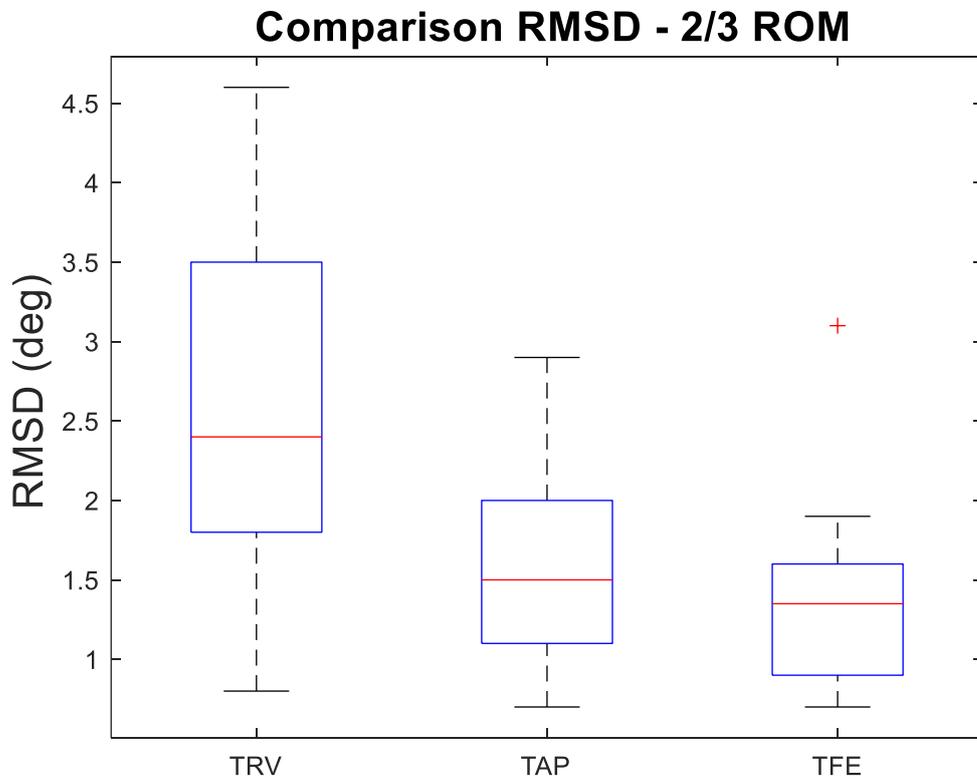


Figure 5.4: Boxplot of exercises involving trunk - medium amplitude.

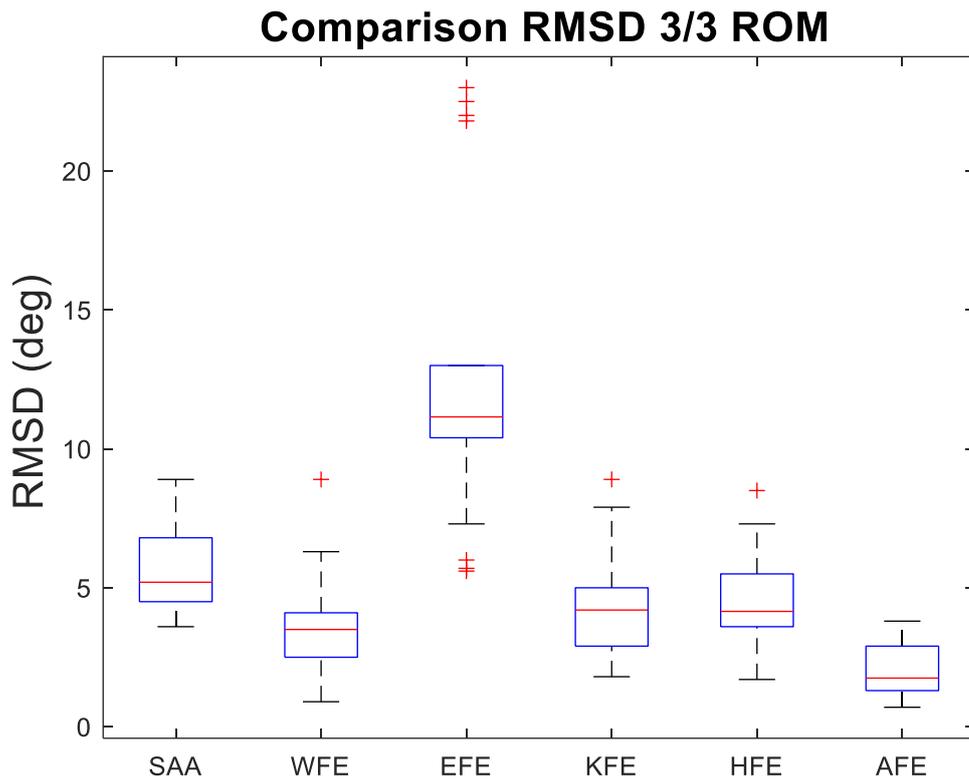


Figure 5.5: Boxplot of exercises involving upper and lower limbs - large amplitude.

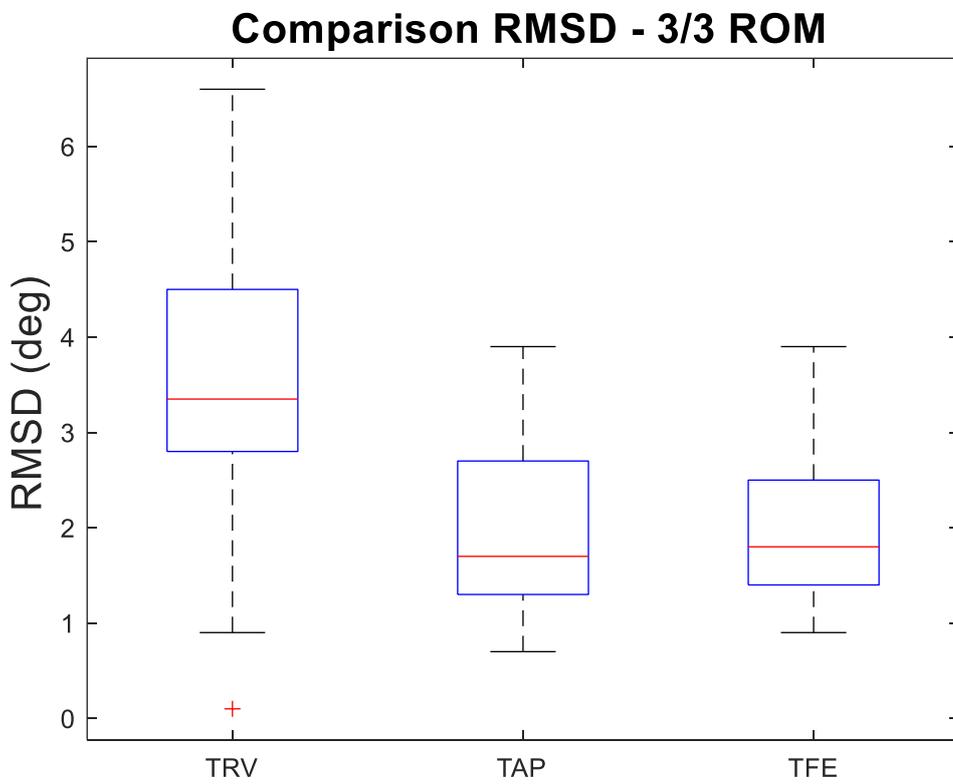


Figure 5.6: Boxplot of exercises involving trunk - large amplitude.

Now about the reliability outcomes the ICC results obtained with both systems IMU, and SP were reported.

The colours in the Table 16 were chosen to identify better the different level of reliability: green indicates an excellent reliability, the red a poor reliability and the black moderate and good reliability.

Table 16: ICC values for both IMU and SP.

	IMU	SP
SAA	0.69	0.50
WFE	0.81	0.69
EFE	0.64	0.71
KFE	0.79	0.81
HFE	0.73	0.92
AFE	0.82	0.87
TRV	0.40	0.57
TAP	0.89	0.84
TFE	0.70	0.70

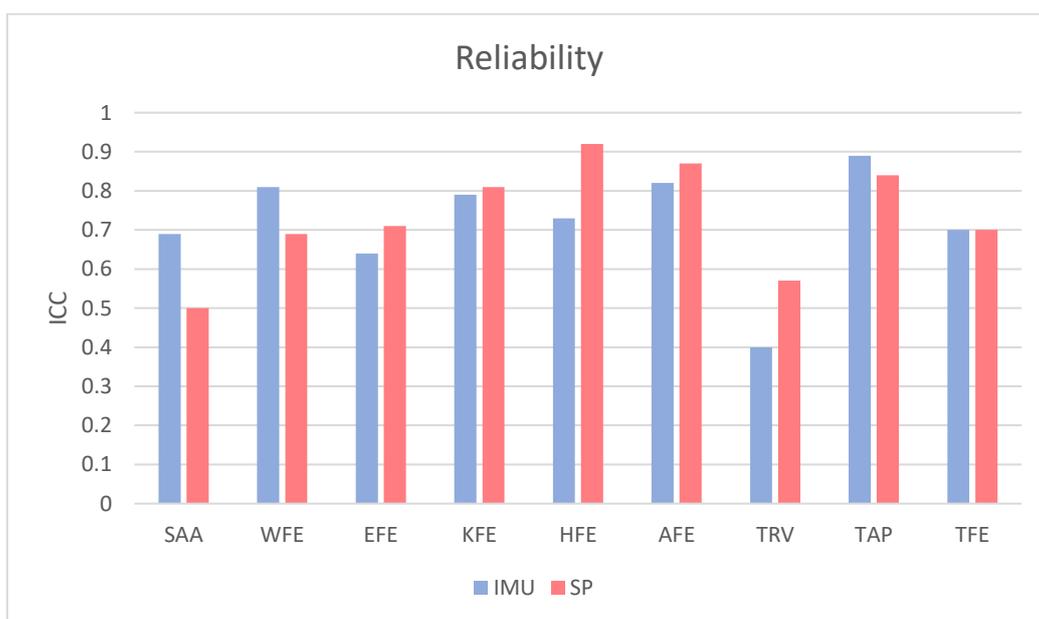


Figure 5.7: Graphic representation of ICC's difference between two systems.

6 Discussion

Inertial sensors are becoming increasingly popular to evaluation of the joint kinematics because they are time efficient and are not affected to all constraints of an optoelectronic system, even if it is still the gold standard for movement analysis.

The most important limitation of SP is the lack of portability, the experiments must be taken into an equipped laboratory and not out of lab.

This study had two principal aims, the first aim was to compare the IMU and SP joint kinematics estimation to test the accuracy of IMU-based system i.e., DoMoMEA by means of RMSD calculation and the second one was the reliability assessment of the ROM extracted from IMU with the calculation of ICC.

About the accuracy the results demonstrated overall to be encouraging with few exceptions. Starting with the shoulder RMSD's average seemed that the error increased with the complexity of movement and in fact for a simple abduction/adduction performed keeping the movement amplitude at 1/3 of ROM max the error was less of 3 deg and this means that the two systems showed a better concordance, but the values reached 5.7 deg during the movement at the large amplitude. During the experiments it was noted that the elastic strap on the thorax shifted a lot especially in the female during the third series of the movement when the arm lifted and tried to reach 3/3 of the maximum ROM. This could have caused some movement artifacts and so a higher value of error.

However, the errors are consistent with those found in literature. For example, in the work of Marrow et al., 2017 [31] was found an RMS error between IMU and a standard lab-based motion capture system for shoulder elevation of about $6.8 \text{ deg} \pm 2.7 \text{ deg}$ (mean \pm SD). The study cited explained this result taking into account the importance of the IMU alignment to the specific segment and affirmed that it was this alignment which permitted to obtain a high level of accuracy.

The most problematic exercise resulted to be the elbow flex/extension with an RMSD until 12.8 deg. A first consideration about this high value could be the different definition of the upper arm axes. Moreover, it is known that the elbow is a complex joint which involves two body segments; arm and forearm where the IMUs were placed. Generally, the forearm is a particularly difficult segment to position an IMU sensor correctly because of the movement of pronation/supination that could rotate the flexion/extension axis of the IMU and in this way it is no longer aligned with the anatomical elbow flexion/extension axis [31].

The RMSD values of other exercises performed highlighted a good accuracy of the angle time series, in fact for wrist, knee, hip, ankle and trunk the error calculations amounted to less of five degrees. This means that gave the results very close to those of the reference system presenting.

The second aim of this thesis was the reliability assessment of IMUs using the ICC. The results obtained were reported in the Table 16, so it is possible to have an overview about trials more reliable than others. Taking as a reference the reliability intervals defined in [23] and reported in Table 6 we obtained a good IMU reliability for wrist, knee, ankle, and the trunk rotation around antero-posterior axis, in these cases the ICCs amounted to 0.79 for knee and reached a value of 0.89 in trunk rotation AP. On the contrary the calculations demonstrated a moderate reliability for elbow, shoulder, hip, and trunk FE.

The elbow, shoulder and hip could be resulted moderately reliable for the elastic strap movement caused by muscle contractions, wobbling of soft tissues, and skin stretching/sliding [32]. In fact, the skin motion artifact is a common source of error to all body mounted devices, especially for the thigh sensor that was been the one more susceptible to skin and soft tissue artifact because the segment where it had to be placed is surrounded by a substantial amount of soft tissue [33].

Moreover, the movement with a very low ICC and so with a poor reliability is the trunk rotation around vertical axis, the value amounted to 0.40.

A possible reason could be found just in the requested trunk movement that was weakly constrained and its amplitude was difficult to reproduce between sessions. A similar result was observed also for the reference, but in this case to obtain the reliable measurements of trunk motion need to pay attention on other factors, for example a more specific definition of bony landmarks to be able to place marker in the correct point of interest. In some subjects this resulted very difficult because could not palpate the point on subjects with a major amount of adipose tissue [34].

Overall, the ICC results highlighted a good agreement between the IMU and reference measurements but for the wrist, hip, and trunk rotation around the vertical axis.

In particular, the hip reliability computed with SP resulted excellent, this because for this joint probably has been easier to replace the marker correctly in the same position between two sessions.

7 Conclusions

The results of this study showed that the proposed technique could provide valuable kinematic data because was obtained acceptable accuracy of commercially available IMUs, i.e., the instruments used in DoMoMEA for measuring upper and lower body kinematics compared to a standard marker-based motion capture system.

The overall experiments demonstrated good results about accuracy but even higher accuracy in position or angle estimation is necessary especially for shoulder and elbow.

One possible solution is to apply further constraints on the physiological models of joints to improve estimation accuracy.

Such *Zhou,2010* [35] suggested for example, further constraints on the elbow angle could lead to improvements of the estimation of the relative orientation of the upper arm and forearm.

The accuracy and the reliability were studied to have a more complete analysis of the IMU-based system, that has a lot of advantages, but it cannot yet replace the gold standard.

In addition, it was important to assess reliability because sensor relocation is an issue that might result in a large error because of the effects of different anatomical characteristics [35].

However, this original technique might open new horizons leading to a better understanding of patients' recovery and in addition to be more and more helpful to clinicians to adapt the therapy to each patient. Inertial sensors would be an excellent alternative to avoid using marker and so, their occlusion. Marker occlusion represents in fact the main disadvantage of the SP and leads to having to reconstruct much of the data acquired by the SP by the interpolation.

On the other hand, IMU measurement allows continuous recording of the joint angle without reconstruction of data.

8 Limitations and future work

The purpose of this study was the validation of a home-based telerehabilitation system, which used the inertial measurements units to compute joint kinematics and give real-time feedback to the post-stroke patients and monitoring their progresses.

These first preliminary analysis highlighted in addition to the advantages also some disadvantages for example in IMUs used the absence of the magnetometer it was fundamental to avoid unexpected ferromagnetic disturbances. But this choice has required the development of joint kinematics algorithms for the positioning of IMU on the subjects' skin in a predefined position.

However, this solution had some limitations such as the round surfaces which did not allow the IMUs to be accurately aligned along the segment axes. Furthermore, the assumption of the sensor-to-segment alignment to be time invariant might not be always true due to the soft tissue artifacts which can be different among the patients.

About the results, instead this work obtained preliminary good results, but it will be necessary to perform other experimental sessions, probably with a greater number of volunteers healthy and also post stroke to be able to give more detailed information about this technology and in particular for the most problematic outcomes, the elbow and shoulder accuracy or the elbow and trunk rotation V reliability.

Especially for reliability the quantity of subjects is relevant, in fact in literature [27] [28] [30] were involved respectively 10, 13 and 23 subjects.

Even the researcher of [36] claim that for calculation of a good reliability the right sample size should be at least 50 and more.

Appendix: the analysis of variance (ANOVA)

To compute the ICC described in (4.2.1), it is necessary to perform the analysis of variance (ANOVA). In this appendix, a theoretical background is given to explain how to compute the MSR, MSC, and MSE terms. The ANOVA is a set of statistical techniques which are part of inferential statistics and allows the comparison between two or more groups of data by calculating and comparing the within-group variability with the between-group variability. The ANOVA checks the null hypothesis which prescribes that the data in all groups must have the same origin, i.e., a stochastic distribution, and the observed differences between groups are due only to chance. This type of statistic methods is the most used in medical research to analyse data obtained from complex experimental designs [37]. The ANOVA is known for its many models which are distinguished from each other by the number of dependent (which can be only quantitative) and independent variables (which can be either qualitative or quantitative). In this study the dependent variable was the average of ROM values, and the independent variable was the time point.

The most popular ANOVA models are:

- one-way model with one independent variable
- Two-way model with two independent variables
- The univariate ANOVA with only one dependent variable
- The multivariate ANOVA i.e., MANOVA which provides two or more dependent variables

A summary about the characteristics of forms of ANOVA is showed in the table below, taken from *Gaddis, 1998*.

Table 17: An overview of ANOVA models

Method	N° Independent variables	N° Dependent variables	Comments
1-way ANOVA	1	1	Simplest form of ANOVA; ≥ 3 groups are defined by a single independent variable
2-way or n-way ANOVA	≥ 2	1	Two or more independent variables are tested simultaneously for their effects on the dependent variable.
Repeated-measured ANOVA	≥ 1	1	Data are obtained from subjects more than once, either across or within treatment groups
MANOVA	≥ 1	≥ 2	Simultaneously tests ≥ 2 related dependent variables

In this study the univariate analysis of variance has been used and it includes different experimental designs among:

- "Between-subjects" design, also known as "independent groups," corresponds to the design in which each treatment or experimental condition is applied to a different group of subjects. When a subject was exposed to one condition he was not exposed to any other condition.
- Factorial or multi-way designs where are examined the effects of two or more independent variables on the dependent factor.
- "Within-subjects" design where the same subjects are tested in different occasions, this type is also called the "repeated measures" design.

Before starting with the application of the technique it is necessary to introduce some terminologies used:

- *grand mean* that is the mean of sample means or the mean of all observations combined, irrespective of the sample, its computation is shown in the expression (1)

$$\bar{\mu}_{..} = \frac{\sum_i \sum_j \mu_{i,j}}{n} \quad (1)$$

where n is the total number of observations.

- *Between group variability* which is the difference between individual averages of each sample and the grand mean. It can be explained clearer with a figure, considering the distributions of Figure A.1 if these overlap the difference is not significant, on the contrary if the distributions don't overlap.

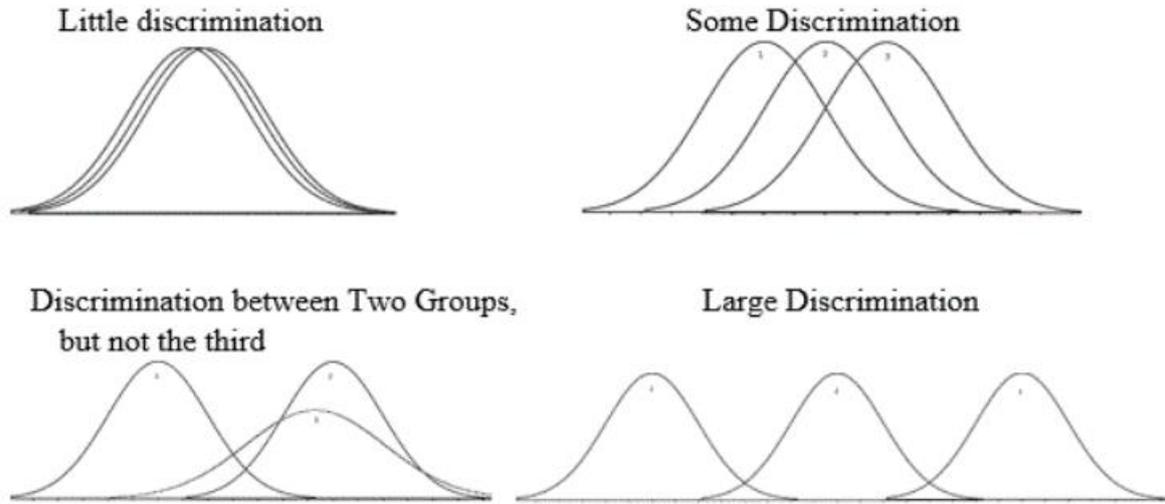


Figure A.1: Some examples of distributions with a different level of discrimination (Source: Psychstat – Missouri State).

To compute this variability is performed the deviance i.e., the sum of the squared difference between the mean of each sample so of each subject i ($\bar{\mu}_i$) and the grand mean ($\bar{\mu}_{..}$) and divided for the degrees of freedom ($k-1$) where k is the number of observations to obtain the means square between control groups like the equations below explain:

$$SS_R = \sum_{i,j} (\bar{\mu}_i - \bar{\mu}_{..})^2 \quad (2)$$

$$MS_R = SS_R / k - 1 \quad (3)$$

- *within group variability* which is the examination of the variability of each point without considering the interactions between samples. To obtain this quantity it will be observed how much each value in each sample ($\bar{\mu}_{ij}$) differs from its respective sample mean. The formulas for the sum of squared and next for the mean square are showed below:

$$SS_W = \sum_{i,j} (\bar{\mu}_{ij} - \bar{\mu}_i)^2 \quad (4)$$

$$MS_W = SS_W / k(n - 1) \quad (5)$$

Due to the experimental design of the DoMoMEA validation, this work relies on the ANOVA repeated measures model which is appropriate for simple or complex designs in which the subject “serves as his own control” or participates in multiple measurements over time [37]. For this type of ANOVA, it is necessary to introduce other deviance formulas; the deviance between subjects (6) and total deviance (8) and obtain the respective mean square to use in ICC formula to be chosen to assess the reliability.

$$SSc = \sum_{ij} (\bar{\mu}_j - \bar{\mu}_{..})^2 \quad (6)$$

$$MSc = SSc / (n - 1) \quad (7)$$

$$SS_T = \sum_{ij} (\mu_{ij} - \bar{\mu}_{..})^2 \quad (8)$$

$$MS_T = SS_T / (nk - 1) \quad (9)$$

In addition, the degrees of freedom (*df*) mentioned are for definition numbers, usually positive integers, that are used to be able to verify some hypotheses about the population from which a sample was extracted and correspond to the number of independent information that are free to vary in the calculation of a given parameter estimation. If the parameter to be estimate is a mean, there is always a limit to the freedom to change the observations [38].

In fact, every time that an analysis is based on the estimation of an average, the number of degrees of freedom will be reduced by one such as is possible to note in the formulas of MS_R , MS_W , MS_C and MS_T where the all the *df* on the denominator are reduced by the unity.

References

- [1] A. O. Mendivil, «Brainstem Stroke: Anatomy, Clinical and Radiological Findings,» *Seminars in Ultrasound, CT and MRI*, vol. 34, n. 2, pp. 131-141, 2013.
- [2] R. A. Grysiewicz, «Epidemiology of Ischemic and Hemorrhagic Stroke: Incidence, Prevalence, Mortality, and Risk Factors,» *Neurologic Clinics*, vol. 26, n. 4, pp. 871-895, 2008.
- [3] E. Gusai, «Development of a home-based neuromotor telerehabilitation system for mildly impaired patients with stroke: the DoMoMEA Project,» *JHBI*, p. 9, 2022.
- [4] B. DM, «Human factors in the development and implementation of telerehabilitation systems,» *Journal of telemedicine and telecare*, vol. 14, n. 2, pp. 55-58, 2008.
- [5] L. M. B. David M Brennan, «Human factors in the development and implementation of telerehabilitation systems,» *Journal of telemedicine and telecare*, vol. 14, n. 2, pp. 55-58, 2008.
- [6] A. Zedda, «DoMoMEA: a Home-Based Telerehabilitation System for Stroke Patients,» in *Engineering in Medicine and Biology Society*, Montreal, 2020.
- [7] E. R. G. P. A. M. Corina Nüesch, «Measuring joint kinematics of treadmill walking and running: Comparison between an inertial sensor based system and a camera-based system,» *Journal of Biomechanics*, vol. 57, pp. 32-38, 2017.
- [8] M. E.-G. a. J. McNames, «Human Joint Angle Estimation with Inertial Sensors and Validation with A Robot Arm,» *IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING*, vol. 62, n. 7, pp. 1759-1768, 2015.
- [9] A. Cereatti et al., «Accurately measuring human movement using magneto-inertial sensors: techniques and challenges,» in *International Symposium on Inertial Sensors and Systems (ISISS) Proceedings*, Hapuna Beach, HI, USA, 2015.
- [10] B. Fasel, «Drift Reduction for Inertial Sensor Based Orientation and Position Estimation in the Presence of High Dynamic Variability During Competitive Skiing and Daily-Life Walking,» PhD Thesis, Lausanne, 2017.
- [11] A. Cappozzo et al., «Human movement analysis using stereophotogrammetry: Part 1: theoretical background,» *Gait & Posture*, vol. 21, n. 2, pp. 186-196, 2005.

- [12] Camomilla V. et al., «IN-FIELD USE OF WEARABLE MAGNETO-INERTIAL SENSORS FOR SPORTS,» in *33rd International Conference on Biomechanics in Sports*, Poitiers, France, 2015.
- [13] K. M. C. A. Caruso M., «An innovative MIMU-based procedure for the estimate of the knee flexion-extension angle,» in *GNB*, Milan, Italy, 2018.
- [14] H. Luinge, *Inertial sensing of human movement*, Enschede, the Netherlands: Twente University Press, 2002.
- [15] R. B. R.L. Gajdosik, «Clinical Measurement of Range of Motion: Review of Goniometry Emphasizing Reliability and Validity,» *Physical Therapy*, vol. 67, n. 12, p. 1867–1872, 1987.
- [16] S. Diaz et al., «Use of Wearable Sensor Technology in Gait, Balance, and Range of motion analysis,» *Applied Sciences*, vol. 10, p. 234, 2019.
- [17] C. A. C. A. Picerno P., «Joint kinematics estimate using wearable inertial and magnetic sensing modules.,» *Gait and Posture*, vol. 28, n. 4, pp. 588-595, 2008.
- [18] J. L. McGinley et al., «The reliability of three-dimensional kinematic gait measurements: A systematic review,» *Gait & Posture*, vol. 29, pp. 360-369, 2009.
- [19] R. B. Õ. S. T. D. G. J. R. Davis, «A gait analysis data collection and reduction technique.,» *Human Movement Science*, vol. 10, n. 5, pp. 575-587, 1991.
- [20] M. S. A. M. L. D. S. T. K. M. d. C. U. & C. A. Caruso, «Analysis of the Accuracy of Ten Algorithms for Orientation Estimation Using Inertial and Magnetic Sensing under Optimal Conditions: One Size Does Not Fit All,» *Sensors*, vol. 21, n. 7, p. 2543, 2021.
- [21] Vicon, *Plug in Gait Reference Guide*.
- [22] McGraw, «Forming Inferences About Some Intraclass Correlation Coefficients,» *Psychological Methods*, vol. 1, n. 1, pp. 30-46, 1996.
- [23] K. TKK, «A guideline for selecting and reporting intraclass correlation coefficients,» *Journal of Chiropractic Medicine*, vol. 15, n. 15, pp. 155-163, 2016.
- [24] J. P. WEIR, «QUANTIFYING TEST-RETEST RELIABILITY USING THE INTRACLAS CORRELATION COEFFICIENT AND THE SEM,» *Journal of Strength and Conditioning Research*, vol. 19, n. 1, pp. 231-240, 2005.
- [25] Shanshan Qin et al., «Assessing test–retest reliability of patient-reported outcome measures using intraclass correlation coefficients: recommendations for selecting and

- documenting the analytical formula,» *Quality of Life Research*, vol. 28, n. 1, pp. 1029-1033, 2019.
- [26] A. Salarian, «MATLAB central file exchange,» [Online]. Available: Arash Salarian (2022). In <https://www.mathworks.com/matlabcentral/fileexchange/22099-intra-class-correlation-coefficient-icc>. [Consultato il giorno 1 July 2022].
- [27] S. SCHNEIBERG, «Reliability of kinematic measures of functional reaching in children with cerebral palsy,» *DEVELOPMENTAL MEDICINE & CHILD NEUROLOGY*, vol. 52, n. 52, pp. 167-173, 2010.
- [28] D.Laroche, «Test-retest reliability of 3D kinematic gait variables in hip osteoarthritis patients,» *Osteoarthritis and Cartilage*, vol. 19, n. 19, pp. 194-199, 2011.
- [29] B. Juul-Kristensen, «Test-retest reliability of joint position and kinesthetic sense in the elbow of healthy subjects,» *Physiotherapy Theory and Practice*, vol. 24, n. 1, pp. 65-72, 2008.
- [30] L. Cai, «Validity and Reliability of Upper Limb Functional Assessment Using the Microsoft Kinect V2 Sensor,» *Applied Bionics and Biomechanics*, vol. 2019, n. 2019, p. 14 pages, 2019.
- [31] M. Morrow et al., «Validation of Inertial Measurement Units for Upper Body Kinematics,» *Journal of Applied Biomechanics*, vol. 33, n. 3, pp. 227-232, 2017.
- [32] A. Cereatti et al., «Standardization proposal of soft tissue artefact description for data sharing in human motion measurements,» *Journal of Biomechanics*, vol. 62, pp. 5-13, 2017.
- [33] H. Dejnabadi, «Estimation and Visualization of Sagittal Kinematics of Lower Limbs Orientation Using Body-Fixed sensors,» *IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING*, vol. 53, n. 7, 2006.
- [34] Margaret Frost et al., «Reliability of Measuring Trunk Motions in Centimeters,» *East Carolina University*, vol. 62, n. 10.
- [35] H. Z. a. H. Hu, «Reducing Drifts in the Inertial Measurements of Wrist and Elbow Positions,» *TRANSACTIONS ON INSTRUMENTATION AND MEASUREMENT*, vol. 59, n. 3, 2010.

- [36] G. N. J. S. Javali SB, «Effect Of Varying Sample Size In Estimation Of Reliability Coefficients Of Internal Consistency,» *WebmedCentral BIOSTATISTICS*, vol. 2, n. 2, 2011.
- [37] M. L. Gaddis, «Statistical Methodology: IV. Analysis of Variance, Analysis of Covariance, and Multivariate Analysis of Variance,» *ACADEMIC EMERGENCY MEDICINE*, vol. 5, n. 3, pp. 258-265, 1998.
- [38] P. Pozzolo, «La tua statistica,» 27 Marzo 2021. [Online].