FACULDADE DE
CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE D
COIMBRA

João Nascimento

Virtual Personal Trainer For Active Ageing

Thesis submitted to the
University of Coimbra for the degree of
Master in Electrical and Computer Engineering

Supervisor:
Paulo Menezes

Coimbra, 2023

This work was developed in collaboration with:

University of Coimbra

UNIVERSIDADE DE COIMBRA

Department of Electrical and Computer Engineering

deer.uc
DEPARTAMENTO DE ENGENHARIA ELETROTÉCNICA
E DE COMPUTADORES

Institute of Systems and Robotics

INSTITUTE OF SYSTEMS AND ROBOTICS
UNIVERSITY OF COIMBRA

Acknowledgments

Gostaria de agradecer ao meu orientador Professor Doutor Paulo Jorge Carvalho Menezes pela orientação ao longo destes meses a qual não só permitiu fazer esta dissertação, como também me permitiu crescer a nível pessoal e profissional. Quero também agradecer a oportunidade que me foi dada para integrar no IS3L, no qual nunca esperei conhecer as pessoas que conheci, quer seja pelo apoio que me forneceram relativo à dissertação, quer seja pelo apoio pessoal o qual fez me crescer numa forma a qual nunca esperei o fazer em tão pouco tempo. Em especial, queria agradecer ao Vishal Gautam pela ajuda com o sistema de captura de movimento, ao pai do laboratório Bruno Ferreira e ao Gustavo Assunção pela disponibilidade quer seja pessoal ou profissional.

Para o Instituto de Sistemas e Robótica da Universidade de Coimbra (ISR), quero agradecer pela oportunidade, material especializado e condições de trabalho pois sem qualquer um destes, a realização deste trabalho não teria sido possível. Apenas posso mostrar a minha gratidão para com todos os que estiveram nestes 6 anos da minha vida académica, os quais apenas conheci graças à Universidade de Coimbra, em especial os meus amigos André Graça, Luís Cavalaria, Gonçalo Valverde, João Limeiro, Tomás Duarte e Alexandre Chaves que tornaram esta experiência educativa numa experiência a qual nunca irei esquecer. Quer seja pelos momentos de apoio incondicional ou pelos vários momentos que se viveu no Troika, AAC, Queimas e Latadas. Neste momento apenas posso agradecer-vos neste texto, mas quem sabe, um dia se for rico, pago um jantar.

Quero agradecer aos meus amigos de Lagos pelo apoio que me prestaram antes da vida académica e durante o qual em conjunto com o apoio de todas as outras pessoas presentes na minha vida fez ser quem eu sou hoje.

Para a minha família, quero agradecer por todos os sacrifícios que fizeram para me colocar na posição onde estou hoje pois se não fosse vocês, não estaria nesta posição privilegiada. Aos meus pais Carlos Nascimento e Lígia Nascimento, agradeço por todos os momentos de apoio os quais fizeram com que permanecesse em Coimbra pois sem eles não teria experienciado muitas das grandes memorias as quais criei enquanto estive aqui. Agradeço à minha irmã mais velha Alexandra Nascimento por me mostrar como ser uma pessoa

decente e trabalhadora, não por palavras, mas por ações. À minha avó Armanda Rosário a qual sempre disponibilizou tempo para mim, recebeu-me de braços abertos quando ia visitá-la e que adora dizer que era uma das poucas a qual sempre acreditou em mim. Para os restantes familiares, embora não sejam mencionados, agradeço pelo apoio que tenham dado quer seja a mim ou à minha família.

Abstract

Performing physical activity sometimes can be a daunting task as most people will not perform it regularly due to a lack of experience, motivation, or injuries that prevent them. However, exercise has substantial health benefits as it lowers the risk of heart disease and osteoporosis and increases our bodies' overall mobility. Furthermore, with the current state of the world and the development of human-computer interaction technology over the years, it is appropriate to think of new ways to interact with each other besides a real-life scenario. The main focus of this dissertation is the development of a virtual personal trainer that can teach and guide older people on how to perform exercises targeted to sarcopenia, which is a condition associated with the loss of muscle mass that occurs with ageing. The exercises were performed by a certified actor where the movement was captured in Motion Capture technology, creating an entity that can teach the correct form of execution of this type of movement. Additionally, the interaction allows the person to improve their quality of life by teaching them how to perform these exercises while being guided and motivated by the coach, providing a personalized and enjoyable experience. Human pose estimation based on a Machine Learning (ML) solution and a static 2D Grid computed using the user's anatomical points to evaluate performance by comparing it to a coach's performance to provide feedback in real-time so the coach can adapt the communication according to it.

Resumo

A realização de actividade física pode por vezes ser uma tarefa assustadora, uma vez que a maioria das pessoas não a realiza regularmente devido à falta de experiência, motivação, ou lesões que as impeçam. Contudo, o exercício tem benefícios substanciais para a saúde, uma vez que reduz o risco de doenças cardíacas e osteoporose e aumenta a mobilidade global do nosso corpo. Além disso, com o estado actual do mundo e o desenvolvimento da tecnologia de interacção homem-computador ao longo dos anos, é apropriado pensar em novas formas de interagir uns com os outros, para além de um cenário da vida real. O foco principal desta dissertação é o desenvolvimento de um formador pessoal virtual que possa ensinar e orientar as pessoas mais velhas sobre como realizar exercícios dirigidos à sarcopenia, que é uma condição associada à perda de massa muscular que ocorre com o envelhecimento. Os exercícios foram realizados por um actor certificado onde o movimento foi capturado na tecnologia Motion Capture, criando uma entidade que pode ensinar a forma correcta de execução deste tipo de movimento. Além disso, a interacção permite à pessoa melhorar a sua qualidade de vida, ensinando-lhe a realizar estes exercícios ao mesmo tempo que é orientada e motivada pelo treinador, proporcionando uma experiência personalizada e agradável. Estimativa da pose humana com base numa solução de aprendizado de máquina (ML) e numa grelha 2D estática calculada usando os pontos anatômicos do utilizador para avaliar o desempenho ao compararar este com o desempenho de um treinador para fornecer feedback em tempo real para que o treinador possa adaptar a comunicação de acordo com isso.

Contents

List of Figures

List of Tables

List of Acronyms

ISR Institute of Systems and Robotics

UN United Nations

BIA Bioimpedance Analyses

DXA Dual-energy X-ray absorptiometry

MRI Magnetic Resonance Imaging

CT Computed Tomography

SPPB Short Physical Performance Battery

IR InfraRed

ML Machine Learning

CNN Convolutional Neural Networks

DTW Dynamic Time Warping

FastDTW Fast Dynamic Time Warping

DNN Deep Neural Network

VE Virtual Environment

RF Random Forest

MoCap Motion Capture

FK Forward Kinematics

IK Inverse Kinematics

FBX Filmbox, proprietary file format

RGB-D RGB-Depth

EWGSOP European Working Group on Sarcopenia in Older People

EWGSOP2 European Working Group on Sarcopenia in Older People in 2018

MP MediaPipe

1

Introduction

With technological and medical improvements over the decades, life expectancy has increased in Western societies. According to a study made by the United Nations (UN) (2019), [53], about ageing, ”the chance of surviving to age 65 rises from less than 50 percent in the 1890s to more than 90 percent at present in countries with the highest life expectancy”. The highest life expectancy comes with the responsibility to provide systems that successively deal with ageing-related diseases, as these can be highly diverse. One of the main conditions associated with ageing is sarcopenia (from the Greek, ’sarx’ or flesh + ’penia’ or loss), a syndrome characterized by progressive and generalized loss of skeletal muscle mass and strength[12] causing deterioration in strength and physical performance. As this geriatric syndrome progresses, the symptoms increase, leading to more significant vulnerability to accidents, for instance, physical injuries due to falls leading to hospitalization, loss of functional capacity, and poor quality of life, frequently resulting in a risk of death.

Once the development of sarcopenia starts to show clear signs of progression, the suspicion of its presence needs to be evaluated by medical physicians, providing a professional opinion and confirming in this manner the condition. With professional confirmation, the patient can do a series of tests to assess the degree of sarcopenia, including questionnaires and physical exams. In the positive case, the person may then follow specific exercises to reduce or prevent further degradation of the condition since regular specialized exercise is a standard advice from medical professionals to address this condition, showing a tendency to decrease the symptoms. However, adherence to physical activity tends to decrease as people get older, affecting older people with health problems that exercise prevent. It is important to search for the application of strategic promotion of physical activity and, when possible, perform sarcopenia screening tests or other ageing conditions. After the sarcopenia screening tests, depending on the development of the condition, the individual can see their physical capabilities decrease, which negatively impacts the performance in one of the main treatments for the medical condition, physical activity. The degree of sarcopenia severity in terms of physical deterioration can be divided into three categories:

- Pre-Sarcopenia: Loss of muscle mass with no effect in muscle strength
- Sarcopenia: Low muscle mass and low muscle strength
- Severe Sarcopenia: Loss of muscle mass, strength and physical performance

Estimating these parameters (e.g., loss of muscle mass or physical performance) employs a wide range of assessment techniques resourcing physical contraptions or physical activities such as balance tests. In the analysis of muscle mass variables, Bioimpedance Analyses (BIA) can be employed to estimate the volume of fat and lean body mass[51]. Alternatively, for physical performance, Short Physical Performance Battery (SPPB) assesses endurance and strength by observing an individual's equilibrium with different foot configurations. After estimating these variables, patients' conditions can be treated or prevent further aggravation with customized rehabilitation.

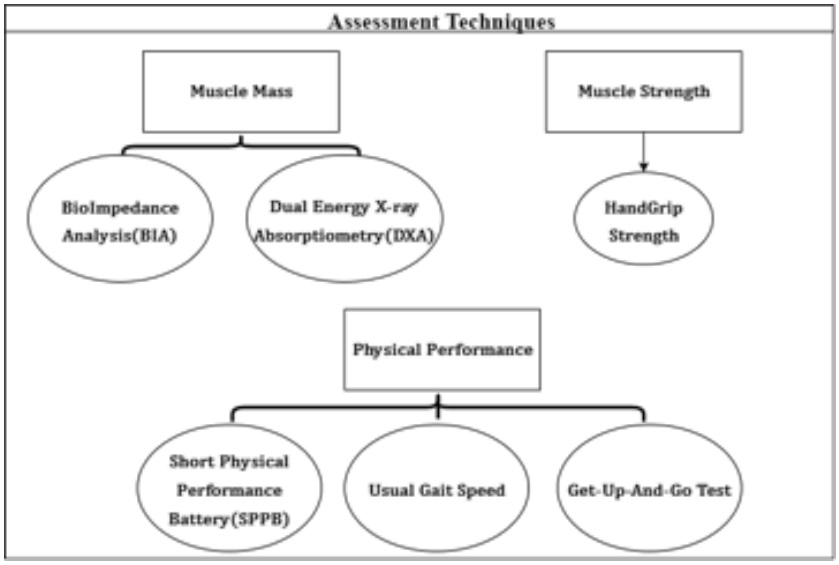


Figure 1.1: Standard clinical practices for the estimation of muscle mass/strength, and function or physical performance parameters[51].

With this knowledge, it is possible to create systems for assessing, treating, or rehabilitating this geriatric syndrome focused on the patient's needs. Although the illness appears in different demographics, older people are the main focus since these can benefit the most from this system, offering personal/professional support for these repetitive and tiring activities, which necessitate a significant amount of intrinsic motivation.

This dissertation focuses on an interactive virtual personal trainer to show and support patients that suffer from sarcopenia and how to prevent by performing specific exercises in an informative and interactive form. The exercises displayed by the personal trainer are animations of certified personal performance as these were recorded through an Optical MoCap system and will serve as a baseline for evaluating the users' performance. The system uses a 2D generalist descriptor to characterize the exercises and therefore create

comparison criteria so the patients' performance can be analyzed. Furthermore, as the patient performs the exercise, the personal trainer will observe their execution in real-time to assist them in case of poor performance or to prevent/aggravate any current injuries.The following section shows this topic's importance by presenting the study's relevance, related work performed to present the current state of the art, and the main objectives to be fulfilled by this work.

1.1 Related Work

The use of interactive Virtual Environment (VE) focused on sports has been conducted to address issues related to physical rehabilitation, lack of physical activity, technique improvement, or weightlifting, which can apply to regular people or high-level athletes.

Physical rehabilitation is a form of assistance for disabled people to treat or at least reduce physical impairments that come with their disabilities. This rehabilitation is helpful for people with lower motor capabilities, such as cerebral palsy[1] patients. One study assessed the possibility of rehabilitating two adolescents with cerebral palsy[10] using a Microsoft Kinect(RGB-D camera)[2] for upper limb rehabilitation to assist people with this condition. The system requires participants to perform regular tasks such as dressing in front of the Kinect sensor, allowing the computer vision algorithms to detect 3D joint points and relevant angles(e.g., shoulder flexion) to the movement. The system, to increase the motivation of the participants to continue rehabilitation, uses audio and video feedback interfaces to provide an engaging application. With this data, a physical therapist could assess whether the results correspond to expectations and provide customized rehabilitation exercises.

Machine Learning (ML) algorithms have an increasing presence in this type of work since they can work with a broad spectrum of data types from RGB or RGB-D frames to sensor outputs to evaluate exercise performance, being one of these examples FitCoach. FitCoach is a virtual fitness coach leveraging robust sensors in users' wrist wearables or arm-mounted smartphones[18] to record signals during workouts that assess performance. The system employs the ML classifier to recognize the anaerobic and aerobic exercises indoors within the first repetitions with a 93% accuracy. The system also provides information such as repetitions per set or calories burned during execution and an exercise score

[1] Cerebral palsy (CP) is a term denoting a group of non-progressive, non-contagious motor conditions that cause physical disability in human development mainly in the various areas of body movement (Rosenbaum et al., 2007).

[2] Kinect is a Microsoft movement sensor that incorporates RGB cameras with InfraRed-rays projector creating an RGB-Depth camera. This type of sensor can detect and map the depth using mathematical methods based on structured lighting and deformation of surfaces. Allowing for the computer vision system to estimate the depth and gather information related to the surfaces of the objects in the scene

based on a reference baseline. The main disadvantages of this system are the need for synchronization of multiple sensors and extraction of the features that require pre-processing of the sensors' signal. One example of these factors is the users' facing direction, which is essential to perform recognition, determined by asking volunteers to perform unrelated physical activities such as running in four different directions. In addition, the user in first-time usage needs to perform particular exercises for exercise recognition and label the corresponding exercise types to obtain the features used in the ML classifier.

As this previous example shows, ML algorithms can provide good results in an exercise recognition task. However, estimating the proximity between a professional's performance and a novice's is also important. For this purpose, Aouaidjia Kamel and Bowen Liu developed iTai-Chi[21], a system that uses Convolutional Neural Networks (CNN) to estimate the learners' pose, being the motion evaluated through a frame-to-frame comparison with the template motion. After exemplifying examples of the Tai-Chi techniques through video format, iTai-Chi records the learners' performance of the martial art through an RGB-D camera for the system assessment against professional motions. The evaluation shows the results to the user to facilitate the correction and location of potential errors in the performance. The main usage of this system is to assist elder Tai Chi practitioners and students without prior knowledge of the martial art to overcome learning obstacles and improve their skills.

This type of work, as shown, is useful for technique improvement but also for weightlifting, which is an activity with many misconceptions about the correct and risk-free way to perform the exercises, creating an opportunity for coaching automation applications. With this concern, Ammar Yasser and Doha Tariq developed Smart Coaching to automate an athlete's coaching using an RGB-Depth Camera to detect misplaced joints of the athlete before an injury can occur[56]. The system uses a fixed IR camera setup to obtain the sets of joint coordinates that characterize the coaches' or athletes' performance. Detection of the correct execution of these exercises occurs using the Fast Dynamic Time Warping (FastDTW)[3] FastDTW classifier to address Deadlift, Squat, and Shoulder Press movements. The main advantage of this classifier when compared to the previous ML algorithms is the ability to classify performances with different quantities of frames(e.g., an athlete may take 3 seconds to complete a lift while another might take 5 seconds) since it classifies according to the sequence of frames[56].

The capability to impact our fitness journey with technology is increasing as this development allows a combination of real-time exercise evaluation using video approaches with VEs. These environments can provide controlled ambients, facilitate stimuli placement,

[3] Analysis technique for time series as Dynamic Time Warping (DTW) is a method that compares and aligns two temporal sequences. The similarity between these can be observed for the same action even if it has differences in time and speed.

and customise personal trainers applications that provide guidelines and assessments while performing actions. For example, a virtual personal trainer using an RGB-D camera for domestic usage captures user actions and compares them with standard actions to create a fitness score representing how well users perform actions in real time[20]. With the standard actions stored in a database, a machine learning algorithm, Adaboost classifier, can process this data to establish a proximity relation between the user action and the standard action classes. After establishing this proximity, a fitness score utilises DTW and 3D Euclidean distance between these actions to provide a final assessment.

Following a methodology similar to the previous system, a virtual trainer with real-time feedback capabilities, assesses and provides a fitness score to different exercise postures using data acquired with a Kinect sensor[24]. The system recognizes exercises using Random Forest (RF) machine-learning algorithm. It provides real-time feedback that relies heavily on comparing the expert's pose and the classification of the user's posture. The feedback is conveyed using a confidence score where the closer the user is to the expert, the more significant the score. The system's average performance is 96% accuracy when applied to 9 exercises for ten individuals.

Apart from sports and rehabilitation, dance teaching has been impacted by motion capture systems by examining techniques using 3D spatial positions in real-time. To address this issue, Yuan Peng developed a system that uses a dataset of 18 dance movements captured subjects' executions through an optical motion capture system[40] as it can identify disparities between poses using a 3D mode's similarity comparison method based on Euclidean norm. However, the system had flaws as it needed users with a MoCap suit for the comparative analysis that was not rigorous enough to consider motions with high amplitude and people with different human body proportions(e.g., height, thinness).

To conclude this topic and explore systems that address issues related to the population of interest, one study stands out regarding the risk of falls. This study developed a visualization system for human balance ability using an improved version of the movement sensor, Kinect two[19]. Human balancing can be seen as a displacement of the body's centre of mass over a support polygon defined using a foot track, which differs between people. When assessing the balance, the centre of mass usually is estimated with anthropometric tables in the biomechanics field or recordings with the support of specialized shoes and a force platform[19] against average population baselines. This work uses a camera to estimate and track real-time variations of the subjects' body mass in the support polygon, providing a more realistic metric of fall risk without the need for physical equipment on the individual. The displacement of the centre of mass was estimated using zero rates of change of angular momentum(ZRAM[4]), providing a projected point on the support poly-

[4]Method presented by Goswami to determine humanoid human balance using the position of the centre

gon and therefore creating a stability index estimated based on the distance between the centre of the polygon and the projected point. The system employs an index representation based on a colour scheme on the skeleton ranging from green to red, providing an accurate and easy-to-understand metric for people without intrinsic knowledge of the concept.

1.2 Main Objectives

After researching the current state of the art regarding virtual personal trainers and development in the performance of assessment, treatment, and rehabilitation of people with sarcopenia, it is clear that there is an opportunity to create an immersive system for this purpose using this type of technology. For this purpose, this dissertation has the following objectives:

- Develop an application that is targeted at elderly population and that addresses the guidance and promotion of physical exercises execution and include support for the assessment of some sarcopenia indicators.
- Contribution to the development of an interactive virtual personal trainer based on research regarding principles of motivation to promote an interactive learning environment.
- Design comparison criteria for evaluating the prescribed specific exercises using a machine learning algorithm for human pose estimation.
- Development of real-time feedback based on the user's monitorization while performing the exercises.

1.3 Dissertation Structure

The structure of the dissertation is the following:

- **Chapter 1** provides the reader purpose of this dissertation by disclosing the relevance of this work and the main objectives defined after the researched relevance of the study.
- **Chapter 2** defines the ageing condition, sarcopenia. The definition, health and financial impacts, evaluation and prevention.
- **Chapter 3** shows the main animation concepts that were followed during the execution of this work.
- **Chapter 4** focuses on the exposure to the reader regarding human pose estimation and its application in this work.

of mass, vector of ground reaction forces and the point of application of the force. The projected point minimizes the rate of change

- **Chapter 5** presents the development of the system using the previous concepts.
- **Chapter 6** defines the application for the pilot study presented in the results chapter.
- **Chapter 7** explains the results, evaluation of the failure cases of the human pose estimator, pilot study and the procedure to analyse the system's usability and the discussion of the results.
- **Chapter 8** defines the conclusion of the work based on the current work.

2

Sarcopenia

Preventing or delaying the onset of physical deterioration is an increasingly important goal because more individuals are living well into their eighth and ninth decades due to the advances in the health medical care industry. So findings ways to prevent or delay physical deterioration in later years has become an important goal for gerontology researchers and practitioners throughout the world[45]. The proposed term ‘sarcopenia’ appeared for the first time in 1989 by Irwin Rosenberg to describe a progressive decline in skeletal muscle mass (MM) related to ageing, and that may lead to decreased strength and functionality[1]. Nowadays, sarcopenia was revised by the European Working Group on Sarcopenia in Older People in 2018 as a muscle disease. Enfacizing low muscle strength as the critical characteristic of sarcopenia. Sarcopenia is associated with a significantly greater risk for poor health outcomes, including disability and functional impairments, increased risk of falls, more extended hospital stays and an increased risk of mortality[31].

In financial terms, sarcopenia is costly to healthcare systems. The presence of sarcopenia increases the risk of hospitalization and the cost of care during hospitalization. An estimation in 2000 showed that the USA spent $18.5 billion in direct healthcare costs related to sarcopenia alone[31]. Among older adults hospitalized, those with sarcopenia on admission were more than 5-fold more likely to have higher hospital costs than those without sarcopenia[13]. To analyze the reasons behind this increase in costs, Ana C. Antunes et al. Daniela A. Ara conducted studies in hospitalized patients revelling that sarcopenia increases the risk of pressure ulcers, infections, autonomy loss, and increased length of hospital stay[1]. With these concerns in mind, preventing, recognizing and screening for sarcopenia and developing steps for its treatment has become a significant public health challenge[31].

2.1 Evaluation

Sarcopenia is a geriatric condition that develops over time, deteriorating muscle mass and functionality; however, if caught early, this is preventable or reversible in some cases. Be-

fore the evaluation for sarcopenia can begin, it is crucial to perform a screening test using certified questionnaires to detect the presence of this condition. The questionnaire needs to extract this information from the user so the system or trained professional can use this information for the well-being of the user. For this purpose, one of the most used methods to assess experiences is questionnaires targetting this specific objective. SARC-F is a certified diagnosis questionnaire developed with the primary goal of rapid assessment for sarcopenia. The European Working Group on Sarcopenia in Older People recommends this test to introduce the assessment and treatment of sarcopenia into clinical practice. The main advantages are being inexpensive, convenient for sarcopenia risk screening and consistent at identifying the risk[13]. SARC-F is a self-reported questionnaire by the patients using five questions:(i) Muscular Strength, (ii) Walking Assistance, (iii) Getting Up From a Chair, (iv) Go Up the Stairs and (v) History of Falls.

Component	Question	Scoring
Strength	How much difficulty do you have in lifting and carrying 10 pounds?	None = 0 Some = 1 A lot or unable = 2
Assistance in Walking	How much difficulty do you have walking across a room?	None = 0 Some = 1 A lot, use aids, or unable = 2
Rise From a Chair	How much difficulty do you have transferring from a chair or bed?	None = 0 Some = 1 A lot or unable without help = 2
Climb Stairs	How much difficulty do you have climbing a flight of 10 stairs?	None = 0 Some = 1 A lot or unable = 2
Falls	How many times have you fallen in the past year?	None = 0 1-3 Falls = 1 4 or more Falls = 2

Table 2.1: Table with the SARC-F's questions and their corresponding options.

Each questionnaire answer provides a score between 0 and 2, whereas the sum gives a scale between 0 and 10. With this score, it is possible to infer the presence of sarcopenia, where 10 is the worst possible case, and 0 is the healthier case. Sarcopenia becomes a concern for the person once the score is equal to or higher than 4. After the evaluation of the user through questionnaires, it is necessary to perform a series of tests to estimate

how the person is affected. There are three types of tests that the patient can perform to assist medical professionals in their assessment. For each type of test, there are several procedures that the user can perform, having in this manner the ability to choose the most comfortable method that will not harm or aggravate their current situation.

After a positive outcome from the SARC-F questionnaire, there are three more parameters to confirm and evaluate the condition's severeness for a more detailed assessment, (i) Muscular Strength, (ii) Muscular Mass, and (iii) Physical Performance.

Muscle strength can be evaluated by measuring grip strength or performing a chair stand test. The grip strength test is the most simple and inexpensive, with a powerful predictor of poor patient outcomes such as longer hospital stays[13], using a calibrated handheld dynamometer and comparing the results with the appropriate reference population. The chair stand test mainly evaluates the strength of the leg muscles, measuring the time for a patient to rise five times from a seated position without using his/her arms. Accordingly to the EWGSOP2, there are several procedures[13] to measure muscle mass/quantity:

- Magnetic Resonance Imaging or Computed Tomography
- Dual-energy X-ray absorptiometry
- Bioimpedance Analyses
- Calf circumference

Magnetic Resonance Imaging (MRI) or Computed Tomography are gold standards for healthcare for non-invasive assessment. However, they are not commonly used due to high equipment costs, lack of portability, demand for highly-trained personnel for usage, and cut-off points for low muscle mass are not well defined. Dual-energy X-ray absorptiometry (DXA) is a more widely available instrument, non-invasive and performed in the clinical and research practice for measuring muscle mass. The disadvantage of this equipment is the lack of portability and consistency when performing measurements with different DXA instrument brands, and the patient's hydration status can influence the measurements[13]. BIA equipment does not directly measure muscle mass. However, it estimates muscle mass, relying on whole-body electrical conductivity and a conversion equation that occurs after a calibration based on a reference DXA-measured lean mass of the specific population[13]. BIA equipment is affordable and portable, with the disadvantage of being influenced by the patient's hydration status. These measurements are prefered over the DXA measurements. Finally, calf circumference has been shown to predict performance and survival in older people[13]. It is mainly used as a diagnostic proxy for older adults when other muscle mass methods are unavailable.

After the muscle mass, the individual needs to perform a physical performance assessment. Physical performance is an objectively measured whole-body function related to locomo-

tion involving muscle and central and peripheral nerve function, such as balance[13]. This assessment needs to consider the individual's current physical condition since he/she can have locomotion issues linked to ageing diseases such as osteoporosis. For this purpose, this assessment can be done using one of the following methods: (i) Get-Up-And-Go Test, (ii) Gait Speed, (iii) Short Physical Performance Battery.

Get-Up-And-Go Test or Timed-Up and Go test (TUG) is a mobility evaluation consisting of getting up from a chair, walking 3 meters and coming back at a comfortable speed returning at the end to the chair and ending the test by sitting back on the chair. In case of failure, the person has severe sarcopenia since the individual will perform this exam after not being able to fulfil the first assessment. Gait speed is a widely used test to analyse walking ability and endurance in practice[13], mainly used in the 4-min usual walking speed test where the gait timing is measured. Although this is one of the best tests to perform, it needs a significant amount of space since the participants need to walk up to 400 meters. SPPB is a composite test that includes the assessment of gait speed, a balance test and a chair stand test. Combining the scores of each test, the final score can reach up to twelve points, whereas a score of up to eight indicates poor physical performance. Although SPPB is used more in research applications, this is the one that provides the most insight into the individual condition.

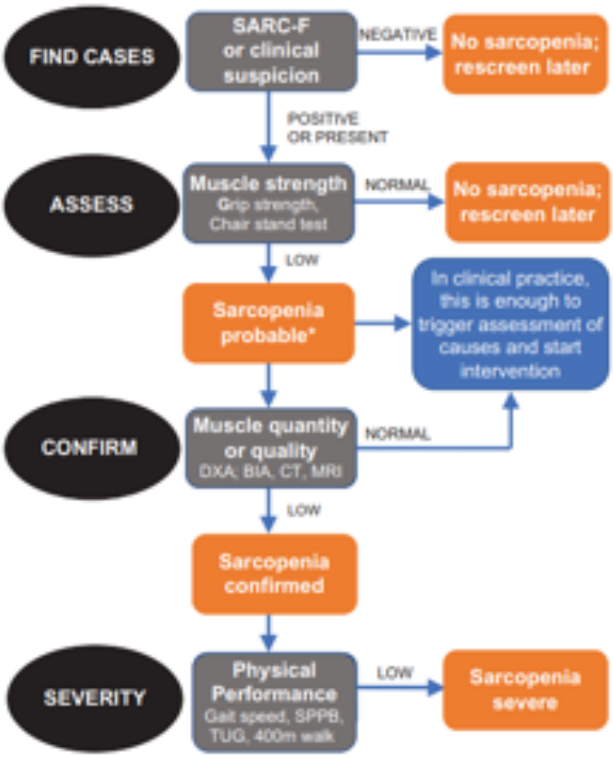


Figure 2.1: Sarcopenia: EWGSOP2 algorithm for case-finding, making a diagnosis and quantifying severity in practice[13].

2.2 Prevention

The following chapter provides guidelines to prevent this medical condition so the individual can perform daily activities and strenuous tasks such as physical exercise, being this the main focus of the dissertation. The referred daily tasks can have different physical demands, ranging from self-care activities such as dressing or bathing to more strenuous tasks such as carrying grocery bags, domestic work, or more advanced activities like physical exercise. Detection of low physical performance predicts adverse outcomes, including falls, fractures, physical disability and mortality[13]. After assessing the performance, it is essential to evaluate two significant components contributing to this worsening: (i) Nutritional Regime and (ii) Muscle Strength Training.

Knowing the two significant components that contribute to the worsening, the first assessment, the person's nutritional stage, is essential since it is one of the primary energy sources for all activities. This assessment is relevant for this geriatric condition since sarcopenia is related to and at least partly over-arched by the general term of malnutrition. However, nutritional assessment has many categories a person can fit into, and Figure 2.12 shows the possible outcomes that a person can be categorized. For this purpose, the European Society of Clinical Nutrition and Metabolism (ESPEN) gathered to identify validated screening tools for subjects at risk of malnutrition[9].

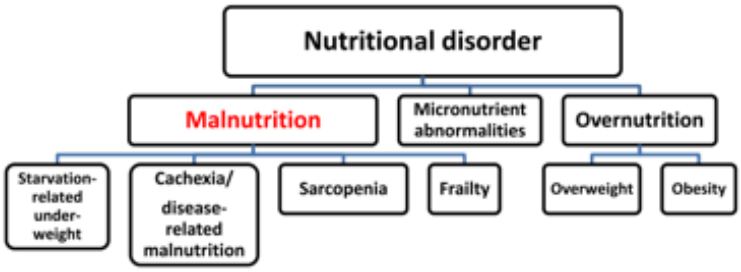


Figure 2.2: Diagram showing all the possible outcomes of the nutritional assessment[9].

ESPN determined that for individuals identified by screening as at risk of malnutrition, the diagnosis of malnutrition should be one of two options. The first option is by the low mass index or BMI (< 22 kg/m^2 for subjects older than 70 years), and the second option is the combined finding of weight loss with the reduced BMI or low fat-free mass index or FFMI (< 15 and < 17 kg/m^2 in females and males, respectively). FFMI is measured using devices such as Bioimpedance Analyses, Dual-energy X-ray absorptiometry, Computed Tomography, ultrasound or magnetic resonance imaging (MRI)[9]. The nutritional assessment is not of concern for this dissertation; however, the documentation is vital since it is one of the main components of the evaluation and, in the negative case, for the prevention of sarcopenia.

Once the nutritional assessment is complete, the second component of the evaluation, muscle strength training, can be addressed. Before the training can begin, it is crucial to evaluate the current condition of the individual since the performance of some types of exercises can aggravate or even create new injuries, which have drastic consequences for older adults. Screening tests for sarcopenia can be a great starting point to assess the current physical condition of the user since the presence of the condition will heavily affect the physical function that the individual has. However, to detect this geriatric condition, it is necessary to perform a series of tests to assess the presence of the condition and, if present, the degree to which it affects the person. The following evaluations define the process of identifying the presence and the category of sarcopenia: (i) Muscular strength to identify Sarcopenia, (ii) Muscular mass to confirm Sarcopenia if diminished muscular strength and (iii) Physical performance to evaluate condition's degree of seriousness.

After the nutritional assessment and the screening test for sarcopenia, it is essential to assess the person's muscle function to provide the best training program possible. There are various training programmes to develop muscle mass and improve overall strength. However, these programmes, most of the time, are general and not customized to the individual, which can have several negative impacts on the physical health of the elderly population. Customized workout routines can be provided by assessing the person's muscle function through questionnaires and performance tests. This evaluation relies primarily on questionnaires such as the Rikli and Jones for older adults.

Question	Can do	Can do with difficulty	Cannot do
Take care of own personal needs	2	1	0
Bathe yourself, using tub or shower	2	1	0
Climb up and down a flight of stairs	2	1	0
Walk outside (1 or 2 blocks)	2	1	0
Do light household chores(cooking or dusting)	2	1	0
Shop for groceries or clothes	2	1	0
Walk 1/2 mile (6-7 blocks)	2	1	0
Walk 1 mile (12-14 blocks)	2	1	0
Lift and carry 10 lb (full bag of groceries)	2	1	0
Lift and carry 25 lb (medium to large suitcase)	2	1	0
Do heavy household activities(vacuuming)	2	1	0
Do strenuous activities(hiking, bicycling)	2	1	0

Table 2.2: Rikli and Jones Questionnaire with the option's score[44].

The Rikli-Jones Questionnaire is an adapted version of a composite physical function (CPF) 12-item scale designed to assess and discriminate across a wide range of functional abilities, ranging from basic activities of daily living (e.g., dressing or bathing) to advanced activities such as strenuous sports/exercise activities[44]. This type of CPF requires participants to answer each of the 12 items with can do (2 points), can do with difficulty (1 point) or cannot do (0 points), resulting in a potential range of scores from 24 (full function) to 0 (unable to perform any of the activities).

With the CPF score, it is possible to provide exercises that are accordingly to the current situation of the person and their current needs in terms of prevention or rehabilitation of this ageing condition. This score can infer the category where the user fits regarding the severity of the condition. When the score is lower than 14 points, the user has low functionality; between 14 and 23 is moderate functionality and above this is elevated functionality which corresponds to the healthier case. Using Otago Exercise Program[39], there is a wide range of exercises for each category, allowing the user to perform safe exercises targeted to the upper and lower limbs of the human body with relevant functionality for daily activities. Roberta Rikli et al. Jessie Jones, the authors of the adapted CPF[44], proved its validity by:

- Determining correlation with previously published scales
- Comparing CPF scores with treadmill performance (standard criterion measure of functional capacity)
- Looking at the sensitivity of the test concerning identifying levels of functional ability

In the ideal case, this scale would be performed in conjunction with a 6-min walk test[44], but for this dissertation, the Rikli-Jones will be used to assess the user's functionality category.

3

Animation Concepts

Human character animation over the years has been an increasing interest in development due to its critical role in entertainment content production such as fiction films, video games or virtual reality. With this interest, the development of tools increased to assist designers and artists in their work's creativity and realism, as in recent works related to digital doubles to bring to life long-time deceased actors[34]. However, to provide a high-quality tool, it is necessary to consider the challenge of human character modulation due to different movements influenced by various factors, for example, mood, intentions or activity[34].

This chapter presents the animation concepts that support the work carried out. These concepts cover the definition and functionality of hierarchical models that form the virtual personal trainer, the animation procedures for these models and the post-processing pipeline to adapt the animation for usage for this dissertation and for most applications of this technology.

3.1 Hierarchical Models

When an animator needs to create 3D complex object representations such as skeletons, the most common approach to this issue is a parent model that uses multiple instances of other simpler models or children where each has its specific affine transformation. Furthermore, with human models, reflected instances such as arms or legs allow the structure to be constructed by reusing parts in multiple positions with specific orientations or scales. In this manner, a skeleton or a complex object can be treated as a list of these instances where each of them has information regarding certain properties (e.g., colour, textures, shaders and relative pose), resulting in a hierarchical relationship between the compound objects and the objects that compose these. Figure 4.1 presents a human hand model to exemplify this relation, showing the finger's degrees of freedom (DoF)[1].

[1] DoF is a variable used to describe the position of that part.

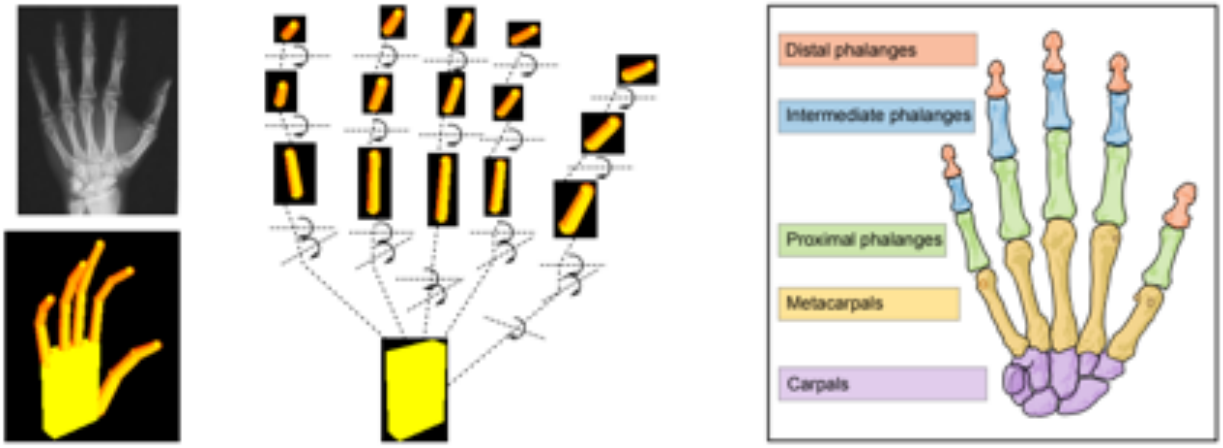


Figure 3.1: Human hand model composition with their respective degrees of freedom (DoF) and the notation for each part.

Using this hierarchical model as a reference, it is possible to see the palm as the origin or root of the model where each of the fingers, acting as children's models, are attached in fixed places, meaning it has fixed translations but with varying orientations or rotations. Each instance can be represented as homogeneous matrices encompassing the rotation and translation in relation to the root or their parent element. It needs to consider their respective DoF and the parent's orientations to compute their orientation. For example, the thumb's proximal phalanx has three rotations or DoF, being its orientation in relation to the palm given by

$$\mathbf{T}_p = \mathbf{T}_{tp,1}\mathbf{R}_{tp,1}\mathbf{R}_{tp,2}\mathbf{R}_{tp,3}$$

Where $T_{tp,1}$ is the thumb translation in relation to the palm and $R_{tp,1}$, i = [1, 2, 3] are the rotations. Relying on hierarchically organized bones, it is possible to employ this methodology in animation. One animation example is Skeleton animation or bone-based animation to animate the interconnected bones that compose a humanoid 3D model and, therefore, the surface representation used to draw the object (e.g., skin, mesh, character)[50]. The hierarchical structure also takes advantage of several controlling techniques, such as forward kinematics, inverse kinematics or keyframing.

3.2 Rigging

After the skeleton definition, rigging is the next stage in an animator's pipeline. Rigging is one of the main stages of the animation process as it allows for the autonomous definition of controllers and manipulators that are later attached to each skeleton body part for the future manipulation of the animator. The constitution of the controllers can range from

simple joints, locators, selection handles, and even an independent graphical user interface (GUI) for control selection[4]. The primary purpose of rigs is to reduce the time consumed on animation by allowing the manipulation of characters using manipulators and controls, which previously was a tedious manual process. However, creating a rig needs to comply with several needs and principles requirements that depend on the character and animation. The following criteria are some of the most important to follow:

- Manipulating the animation controls will not damage or create an unorthodox transformation on the rig
- Simple control structure with the essential amount of controllers required for the animation
- Lightweight and easy to interact rig

Once the rigging artist follows these guidelines, the animator can use the rig to create the character's motion, relying on key concepts such as kinematics and skinning to produce the most realistic outcome.

Kinematics is a subfield of physics that describes the motion of points, objects(e.g., rigid bodies or skeletons), and systems of objects without considering the forces that disturb a stationary state. How controls drive bones to move the mesh can also be defined. This mechanism can be divided into two, forward or inverse kinematics, where an animation, in general, will always have to apply Forward Kinematics or/and Inverse Kinematics to the joints that compose an overall joint chain. Forward Kinematics, usually a default, after the joints are connected one after the other in a chain, applies changes to a child joint based on the rotation of the parent joint being the endpoint of the chain determined by the angles and the relative positions of each joint that it contains. Conversely, Inverse Kinematics have an opposite view, affecting a parent joint based on a child joint. To accomplish this task, it is necessary to give a position in space, work backwards and find a possible way to orientate the joints so that the endpoint of the chain can be placed at that position. This kinematic is helpful in cases where a character needs to touch an object in a selected position.

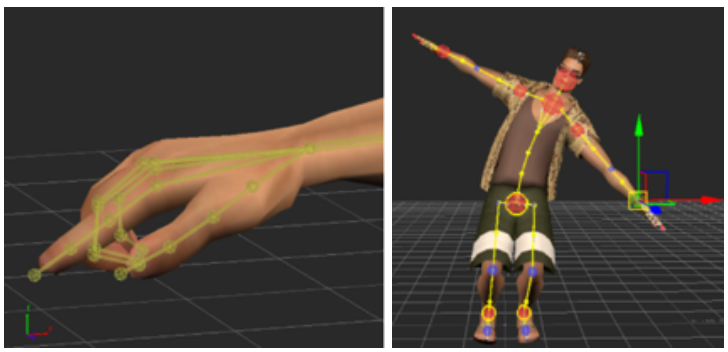


Figure 3.2: On the left side is an FK hand where the finger rotation does not influence the hand and an IK skeleton where the translation of the wrist, affects the whole skeleton[7].

3.3 Skinning

When attempting to produce realistic outcomes, a set of rigid bodies connected via joints is not viable. Skinning is a procedure that addresses this issue by computing the movement of each vertex that constitutes the skin/mesh when the ”supporting” skeleton bones move about their connecting articulations. For this purpose, the models need tissue deformations capabilities similar to the skin when body articulations change their configuration.

At a high level, skinning algorithms can be divided into two categories: Direct and variational methods. Variational methods view the task as an optimization problem, minimizing an objective function, being mainly iterative solvers. On the other hand, direct methods compute the deformations using closed expressions (e.g., without numerical optimization)[22]. When comparing these methods, direct methods are particularly attractive due to their capability to leverage parallel computing, allowing their usage in real-time applications and GPU implementations. After the skeleton bones definition and their respective transformation matrices, it is necessary to define the bones indexes that will influence each vertex with the corresponding weights to apply these methods. Due to graphics hardware considerations, it is common to use at most four non-zero weights for every vertex[22].

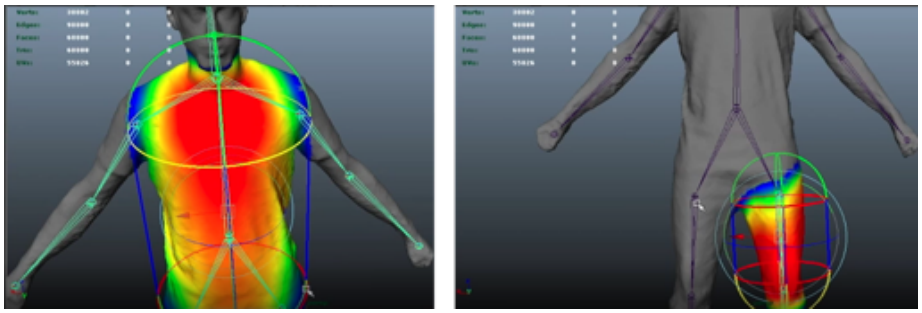


Figure 3.3: Skinning example with a visual colour scheme representing the selected bone’s influence on the mesh vertexes. This influence colour scheme ranges from red (maximum) to green (minimum).

As previously described, bone transformations describe the changes to a resting pose (e.g., T-Pose or A-Pose). More specifically, the skeleton transformations represent the rotations on the articulation joints and, if necessary, other affine transformations. The simplest form to apply the skeleton transformations to the mesh is using the ”Linear Blending Skinning” method. This type of skinning needs the following inputs[22]:

- **Rest pose shape**: typically represented as a polygon mesh, assuming the mesh connectivity is constant(only vertex positions change during deformations), the rest-pose vertices are $v_1,...,v_n \in \mathbb{R}^3$

- **Bone transformations**: Using matrices T_1,..., $T_m \in \mathbb{R}^{3x4}$ to represent spatial transformations aligning the rest pose with its current pose, the input that varies during the animation.
- **Skinning weights**: For a vertex v_i, the weights $w_{i,1}$, ..., $w_{i,m} \in \mathbb{R}$. Each weight $w_{i,j}$ describes the amount of bone j on vertex i. A common requirement is that $w_{i,1} > 0$ and $w_{i,1}$ + ... + $w_{i,m}$ = 1

Using these parameters, linear blend skinning obtains the deformed vertex positions $v_i^{'}$ with the following equation

$$\mathbf{v}_i' = \sum_{j=1}^{m} w_{i,j}\mathbf{T}_j\mathbf{v}_i = (\sum_{j=1}^{m} w_{i,j}\mathbf{T}_j)v_i$$

Skinning is an ongoing field of research in which Linear blend skinning is only one of many approaches to this complex mathematical problem. To address this issue in this dissertation, the solution that has been found is to employ secondary software from Autodesk that employs several of the state-of-the-art techniques that can be found in the references[22, 27].

3.4 Keyframing based Animation

When the animation plays, the viewer cannot distinguish between pictures or frames, seeing a single picture and creating the illusion of continuous motion. Traditionally, the animators would draw or paint individual pictures for the most important actions in a sequence where the action between these key points would be filled with other pictures where this process of filling can be referred to as in-betweening or tweening[6]. However, increasingly demanding realistic animations do not allow animators to perform this task manually due to time and complexity constraints. Nowadays, leveraging the increasing computational load, computer-generated techniques such as keyframing or Motion Capture are preferred since they can perform this task in a fraction of the time and complexity.

Keyframing is an animation technique that uses essential frames or keyframes to define an action's start and end point, letting software perform the task of interpolating between these frames. This type of animation is an improvement compared to previous techniques since it facilitates the animator's work and allows for the economic management of data since it is not necessary to store every frame to define an animation. Apart from this, when performing the computational task of deformation of the mesh/skin given the skeleton configuration (skinning), when it comes to realistic movements, it is hard to define them using explicit parametric equations.

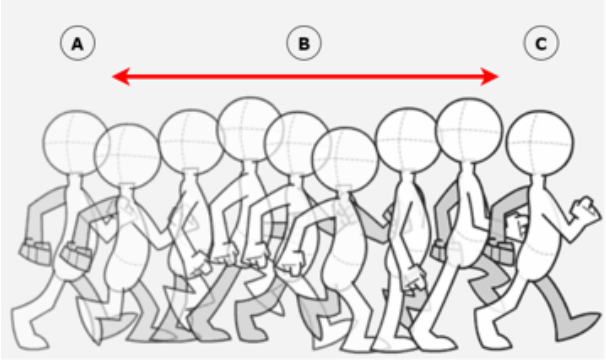


Figure 3.4: Adapted illustration with an animation using two keyframes at the start (A) and end (C). Where the frames in between (B) were computer-generated[50].

Keyframing and Motion Capture based animation are an alternative to defining the ”key” configurations of the skeleton with an associated timestamp. When performing keyframing interpolation on animation, it is necessary to identify the following factors:

- Number of Keyframes on the animation: N
- Animation's Starting instant: t_0
- Duration of the Animation: T

With these factors, it is necessary to compute the fraction of time that exceeds an integer number of complete sequences, for example using a floor function as follows

$$t_f = t - t_0 - [\frac{t - t_0}{T}]T$$

With this variable, the interpolation factor between the keyframes k and k+1 can be determined by

$$\delta = \frac{t_f - t_k}{t_{k+1} - t_k}$$

Afterwards it is possible to compute the interpolated transformation of two translations and/or scale transformations as

$$\mathbf{T}_i = \mathbf{T}_k + \delta(\mathbf{T}_{k+1} - \mathbf{T}_k)$$

However, this method can not be applied to rotation matrices since it would not represent a rotation. An interpolation method between two 3D rotations is SLERP(Spherical Linear Interpolation), introduced by Ken Shoemake in 1985. The main advantage of this approach is the usage of quaternions. For example, a 3x3 matrix contains 9 DoF, requiring

6 constraints to represent rotations. On the other hand, quaternions only need 4 DoF and therefore require only one constraint to represent rotations[22].

Relying on this principle, it is necessary to use the quaternions q_k, q_{k+1} to represent initial and final orientations and spherical linear interpolation (SLERP), defined as

$$slerp(q_k,q_{k+1},t) = q_k(q_k^{-1}q_{k+1})^t$$

or

$$slerp(q_k,q_{k+1},t) = \frac{sin((1-t)\theta)}{sin(\theta)}q_k + \frac{sin(t\theta)}{sin(\theta)}q_{k+1}$$

Assuming the angle between the quaternions $\theta, cos(\theta) = q_k \cdot q_{k+1}$.

3.5 Motion Capture

Another form of skeleton animation that concerns this dissertation is Motion Capture. Using the concept mentioned previously, this type of animation is a method that captures the movements of objects in the real world and then inserts the captured movement data in a tridimensional model that, in some cases, is in a Virtual Environment[36], commonly used in gaming or movies industries. MoCap is widely used in all types of fields involving realistic and precise movement, whether in sports when assessing the performance of athletes[42] or in animation. An example of the MoCap application in the gaming and movie industry is the Lord of the Rings: The Return of the King[36], which used mocap systems that rely on optical motion capture with passive markers to provide the animations for the movie. One of the main requirements for the game version was the synchronization between the original animations from the films and the game's animation on intense scenes such as combat scenes. The fact that the animations were made using motion capture saved countless hours for the production crew since the animations produced by the motion capture were not only used in the film but could be transferred to the characters in the video game.

Figure 3.5: Examples of MoCap application in the movie and gaming industry.

With the need to represent complex animations and reduce animation time consumed, Max Fleischer 1919 invented Motion Capture using rotoscoping to address these needs. Rotoscoping was a manual method which allowed the animator to trace every single frame from live-action footage to create moving images in the style of cartoon animation[28]. Nowadays, rotoscoping employs a digital camera to capture live action and computers to trace the footage, these accomplished using 2D image processing and drawing tool. The first attempt at creating MoCap systems was mechanical systems were the first attempt, usually described as providing limited freedom for movement due to a high amount of cables and prohibitive suits, damaging the capability of producing complicated actions. Eventually, a broad spectrum of MoCap technologies emerged with engineering innovation, ranging from mechanical, acoustical, magnetic, and optical, as the latter can be subdivided into marker or markerless systems.

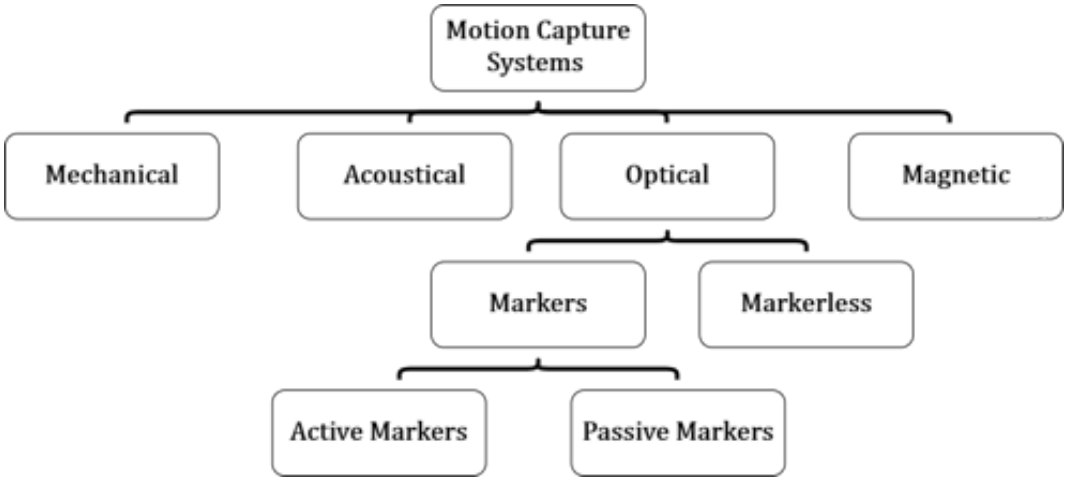


Figure 3.6: A diagram exhibiting all types of motion captures systems as these differ from the design to the core of the functionality itself.

3.5.1 Choice of Motion Capture Technology

As presented above, many technologies are available to capture the movements that define a specific exercise. The advantages and disadvantages of each of these technologies will be briefly introduced to justify the choice.

Mechanical or active tagging systems use potentiometers and sliders on the desired articulations, enabling the exhibition of the position. The advantages of this approach are easily adaptability to stop-motion[2] due to the similarity of functionality and are unaffected by

[2]Stop Motion or frame-by-frame is an animation technique that uses a camera or computer to capture the movement of physical entities. It works the same way as graphic animation, showing frames and waiting for the viewer's brain to do the interpolation. The models are moved and photographed frame by frame. These frames are later mounted on a film, creating the impression of movement.

magnetic fields or unwanted reflections, leading to a short recalibration process. Disadvantages include restrictions on actions due to the usage of cables and prohibitive suits, inability to deal with simple motions(e.g., jumping) since the technology has no sense of ground and recurrent need for recalibration[43].

Another marker-based motion technology is acoustical systems based on a set of sound transmitters or emitters placed on actors' main articulations with receptors positioned on strategic points in the capture room[51]. Emitters provide sets of frequencies sequentially to the receptors and, using the velocity of transmission of sound and triangulation methods between emitters and each receptor, compute 3D positions. With this functionality, the system offers resilience to occlusions or interactions with metallic objects, which affect significant other types of capturing. Nevertheless, the system has drawbacks associated with missing data description in specific instants due to sequential firing[51], limitation on the number of transmitters damaging the quality of animation, and restrictions on movement caused by cable usage that connects the actor's articulations receptors to the antenna of the system and lack of robustness dealing with sound interferences namely external noise.

Magnetic systems employ a set of receptors on actors' articulations to measure position and rotation relative to an antenna. Compared to other systems, the acquisition and processing of data are performed with high precision[51], being useful for characterising simple motions. Weaknesses of this procedure are the necessity of a significant amount of cables connecting to the antenna and the high sensitivity to magnetic occurrences(e.g., magnetic fields) or even the structure of buildings. One example of this technology is the Polhemus Liberty motion tracking[41]. Polhemus uses a cube with three orthogonal coils to each other, where each represents a coordinate axis (e.g., x, y and z) generating a magnetic field. The actor or object has up to 16 Hall sensors that measure the intensity of the magnetic field. By moving these, it is possible to measure the magnetic field variation and therefore compute the position and orientation of each sensor.

Optical capture differs from the others since it can be divided into markerless and marker-based and the latter into passive or active markers. When systems rely on the markerless optical solution to predict 3D human joints coordinates, it is important to state that physical limitations are minimal given that the actor does not need to wear any equipment being the capturing established by novel computer vision techniques that have been revolutionized over the years by machine learning algorithms[11, 30, 8, 48]. The downside of this method is the need for multiple expensive RGB-Depth cameras strategically placed to supply multiple perspectives of the actor and, in this manner, avoid occlusions of body parts which frequently occur in complex animations.

As described above, marker-based optical capture can be divided into passive and active

markers. Passive markers usually are placed in a MoCap suit on the main articulations, captured by high-resolution cameras positioned across the capturing area. Before the actor can perform the movement, there is a need for a calibration phase for unrelated object avoidance purposes, where the actor performs one reference pose to the system(e.g., T-Pose or A-Pose). This reference pose allows the system to emit the IR rays sequentially from each camera to reflect on the markers, allowing the system to synchronize the capture so it can perform better. After the calibration phase, each camera emits the IR rays sequentially onto the markers to supply the system with a 2D perspective of each reflector/marker and with the perspective of at least one more camera, precise reflector 3D coordinates generated with triangulation techniques based on novel computer vision methods. One factor influencing the system's performance is the duration between the emission of the IR rays from different cameras, where the longer duration worsens the system's performance. When a marker is occluded to multiple cameras, the system has to estimate the marker's location from the perspective of the occluded cameras.

After this system's estimation, there will be several possible positions that the marker can be. These are discarded based on the validation of other cameras that can see the marker. Otherwise, the marker will never be defined. Active markers, on the other hand, utilize Inverse Square Law[3] Light Emitting Diode(LED) instead of markers that usually are identified beforehand, emitting their light powered by small batteries[36]. The disadvantages of these techniques are difficulties with occlusions of markers leading to marker swapping, problems of ambiguity since the system relies on a reference pose to start the precise tracking of the markers and dependency for the intended usage of the motion capture data for example, rapid movements need more frames per unit of time and therefore more sampling from the system.

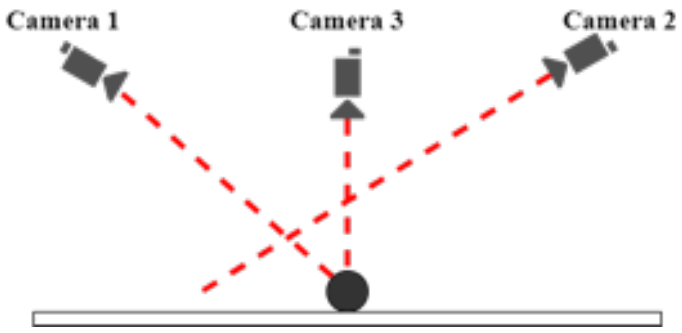


Figure 3.7: Trial and error approach to estimate and validate the marker's position.

With the presented information, one would incline to select the markerless system as the best one due to the improvement over time with the advances in their key points acqui-

[3]Physics formula that describes the intensity of the light to an observer from the source as inversely proportional to the square of the distance from the observer to the source

sition using Machine Learning technology. However, these systems do not provide a critical point estimation compared to the standard marker-based system in terms of performance and precision. The main reason behind this is that marker-based systems have more straightforward processing, which leads to more appropriate rates of acquisition from the cameras, where each of these can be the most precise camera in the market. One example is the cameras produced by Qualisys for marker-based optical motion capture that can function at 1400 frames per second with a 0.06 mm precision. Furthermore, the system can overcome occlusion problems by adding more cameras to more precisely validate the marker's position and markers to form the marker set since these are placed in a well-defined position in the MoCap suit. The relationship between markers allows the system to track and, if missing, estimate the position of the missing markers based on these. These additions can contribute to a higher processing time for the CPU during tracking[51]. Regardless of these factors, this technology had the most advances over the years in motion capture[16, 38, 25, 55], reaching high levels of sampling rate, which enable the capture of fast movements such as martial arts and acrobatics[51]. Also, with the support of ISR, it was possible to access an optical capture room equipped with 10 IR(Infrared) cameras that pulsate Infrared rays that reflect on the passive markers present on the MoCap suit.

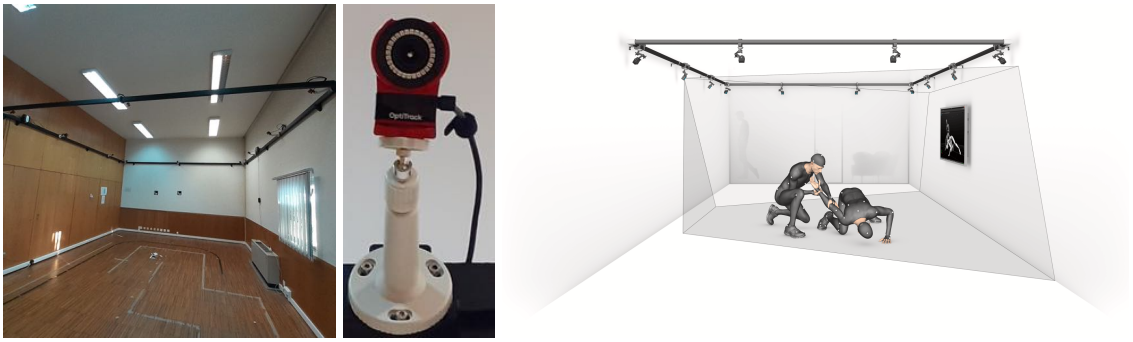


Figure 3.8: The ISR optical motion capture room layout with ten cameras that emit IR rays reflected on the MoCap suit with the passive markers to module the motion.

3.6 Retargetting

One common problem when dealing with animations from several sources is the existence of diverse scales and configurations of skeletons due to application configurations or objectives of the animations. Especially when motion capture data is present in the animation pipeline since one of this technology's main goals is to reuse the information to facilitate development such as video game development and animation production. Weidong Geng et al. Gino Yu provide an overview of the various tasks and techniques involved in the reuse of Motion Capture data[15], providing insight into the procedures employed in MotionBuilder to solve this problem. However, the problem of adapting the motion of one

articulated figure to another with an identical structure but different segment lengths can be solved using retargeting[17]. An example can be the transfer of an animation made by a small child to a skeleton of a tall adult, different in bone length but with an identical structure. Applying the motion to the latter is not a trivial issue, requiring adaptation. The adaptation approach must characterize an activity by retaining high-quality properties using mathematical methods that the system can efficiently compute. High-quality properties are a very broad term since these, in conjunction, define animation's realism. However, to define these in a mathematical approach, it is necessary to define them precisely so the solution can address these. After acquiring the representative motion data, rigging the motion, and the personal trainer, MotionBuilder offers a solution for this retargeting issue using their proprietary software to link these entities. This software considers the different properties that define the skeletons, allowing one to model the actor's action in the most efficient way possible without the loss of realism that optical capturing provides.

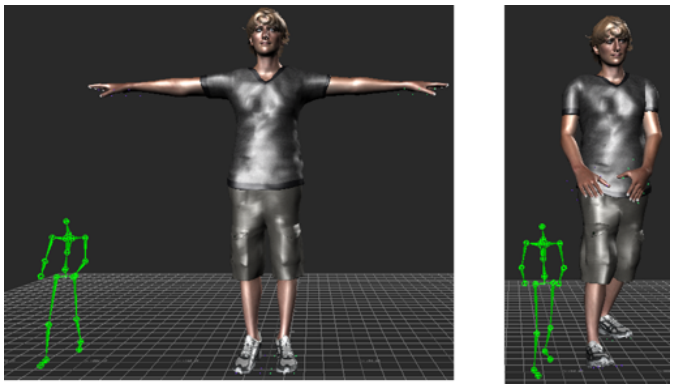


Figure 3.9: Exemplification of the motion capture skeleton represented in green and the 3D humanoid model, being the image on the left side before retargeting and on the right side after retargeting.

3.7 Editing and Exporting Animation

After retargeting and with the animation incorporated into the avatar, editing animation is a critical step in the animation pipeline since the objective of the direction and the animation might differ from the final product that the user will see. One example is using objects such as chairs, weights or exercise balls during the performance of the exercise, which can only be applied after the retargeting step. To accomplish this, MotionBuilder has built-in functions to establish constraints between the skeleton of the coach and the objects. These mathematical methods differ in terms of their objective and overall process. The following constraints examples are some built-in functions of MotionBuilder, where all of the available functions can be seen in the reference[33]:

- (i) Position constraints: The source object constrains the translation of a target object.
- (ii) Rotation constraints: The position of a source object or objects can be used to constrain the rotation of another object.
- (iii) Parent-child constraints: A relationship of parent-to-child between any two objects where the movement of the ”parent” also occurs in the ”child”

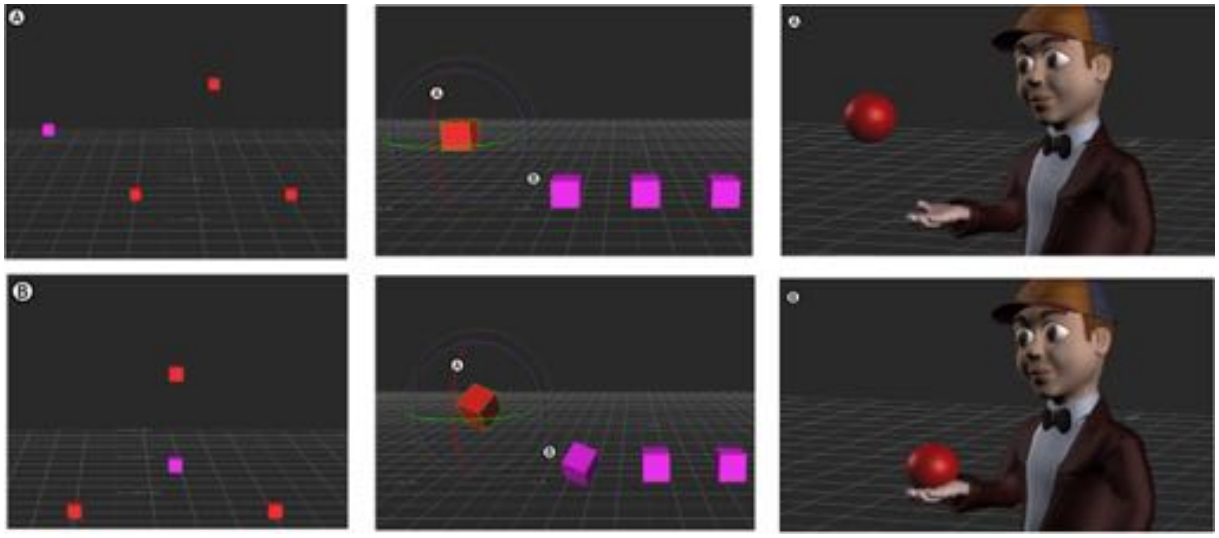


Figure 3.10: On the left side of the figure, the pink marker is constrained to the red markers being only affected when the position of these changes (i). In the middle, the red cube is constrained by the rotation of the source object B (ii). On the right side is a model's hand effector parented to a child-object sphere (iii).

With these valuable tools, editing the animation can modulate the execution of exercises requiring equipment without needing physical equipment with markers during the optical motion capture recording session. Of the previously mentioned constraints, the most used in the animation development was the parent-child constraint due to the simplicity of their usage and the objective(e.g., weights/exercise balls on the hands). Once the objects are in place, the humanoid model needs to be modified to interact in the most natural form possible. Using the IK/FK skeleton from the rigging process and keyframing methods, the avatar's bones' translation and rotation can be adjusted while maintaining human limits, which is highly relevant for this work. This task can range from lower-level undetailed modification of bones such as the spine to high-level meticulous modification of hand bones to provide the viewer with the sensation of genuine interaction with objects. After animation editing, the avatar is ready to be seen by the user and integrated into the system. For this purpose, the exportation of the animation can be in types of formats(e.g., Filmbox, proprietary file format (FBX), BVH, DAE), which will be explained in later chapters as what they represent and the best choice for the dissertation.

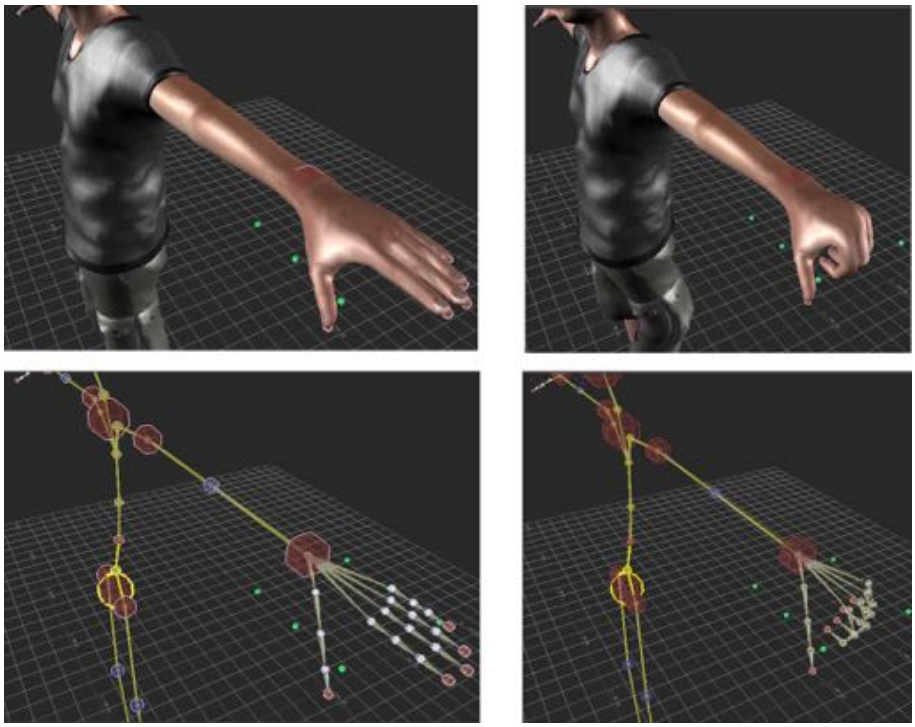


Figure 3.11: Exemplification of animation manipulation using the built-in functions of MotionBuilder.

To conclude this chapter, this dissertation will not explore state-of-the-art approaches used in each of the described topics as the current technology provides the necessary functionality for the objective of the work, being these referenced to contextualize the recent advancement of the technology[34, 26]. To facilitate the reader's understanding of the animation process, the following figure shows, in order, the execution of the fundamental steps described in the topics of this chapter. As it will be described in later chapters, to follow this pipeline, it was necessary to utilize secondary software, proprietary software from Optitrack for the Motion Capture and Autodesk, more specifically, MotionBuilder to adapt the animation.

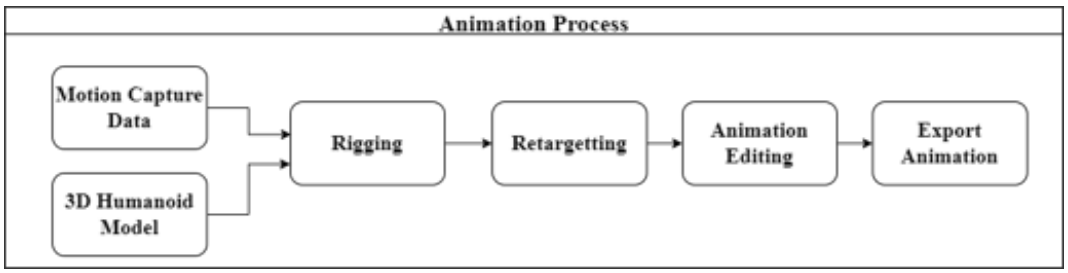


Figure 3.12: Figure with the animation process followed to develop the animations described in the latter chapters.

4

Human Pose Estimation

The detection and tracking of objects or persons is a well-established research topic in computer vision. Its resolution could lead to unique applications in sports, medical feature detection in healthcare or video surveillance. Furthermore, in sports, it is common to utilize this technology nowadays to detect and track players' poses so their actions can be evaluated and improved if needed. In the sports field, this type of evaluation nowadays is used to point out technical errors of high-performance athletes to improve their performance however, this type of evaluation can also be used to assess and detect general errors in a regular person's performance when performing common exercises using the general guidelines on how to perform it.

There are many procedures to detect the presence of a person in the image plane[49]. Nevertheless, one of the most essential distinguishing that can be done is according to the approach taken, which can be single-person or multi-person. The single-person approach identifies a person's pose in the image, the position of this entity, and a set of key points, making it a regression problem. In contrast, the multi-person approach needs to be concerned about other factors like the unknown number and positions of persons within the image; being important that the framework can detect the key points and assemble an unknown number of persons.

Figure 4.1: Pose Estimation using each of the approaches mentioned above, where on the left side is the single-person and on the right side the multi-person approach.

With this knowledge in mind and considering the dissertation's main focus, the exercises'

evaluation will be in real-time on one person's pose at a time, making the single-person approach ideal, simplifying the whole user identification process once the exercise begins. As mentioned before, the single-person approach is mainly a regression problem; however, there are some major differences between the different kinds of methods used to identify or predict the set of key points; being this technique commonly based on one of these frameworks on the keypoint prediction[46]: (i) Heatmap based Framework and (ii) Direct regression-based framework.

The Heatmap framework relies on types of visualization techniques with the main focus on showing the magnitude of the phenomenon as colours using only two dimensions. The colour variation on these heatmaps may have different meanings depending on the application. However, it is useful in tracking methods since it can assign colours depending on the movement of the entities present in the image, facilitating the task of identifying points of interest by reducing the region of interest. This procedure creates heatmaps of all the key points in the provided image, and with the assistance of additional methods, the final stick figure can be constructed. One example of this pipeline is a person detector based on a "stacked hourglass" network, closely related to encoder-decoder architecture, relying on successive pooling steps and upsampling, producing the final set of predictions[46].

On the other hand, Direct regression is characterized by a regressor(e.g., a neural network like a Deep Neural Network) to identify key points from feature maps directly. Where feature maps are a form of a visualization generated by each layer of the neural network to identify different features(e.g., edges, vertical/horizontal lines) that can provide insight into the internal representation of certain inputs. These can be generated using filters, feature detectors to the input image, or other feature maps output of prior layers. With these maps, the regressor can supply an output with the coordinates (x, y) for each keypoint present in the image.

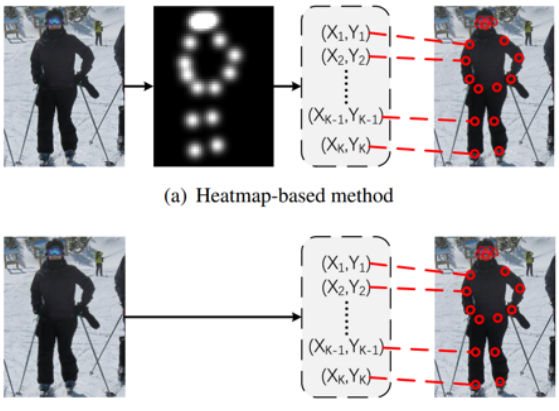


Figure 4.2: Illustration exemplifying the frameworks' pipeline[29].

Regardless of the approach, it is crucial to consider one of the most significant issues this technology faces, occlusions. Occlusions are one of the main problems when performing an accurate pose estimation. Many approaches tried to deal with this issue, whether on multi-person or single-person approaches. These approaches tried novel methods of computer vision and, in many cases, the implementation of ML to assist the task of estimation. The description of these techniques can be seen on the reference [46], using Non-Maximum Suppression to resolve this and eliminate redundant poses or a neural network, Cascaded Pyramid Network (CPN), which included GlobalNet to localize visible key points and RefineNet to handle the ones that were invisible or occluded vital points.

4.1 Selection of Pose Estimator

According to the objective of this thesis, the human pose estimator used needs to be optimized to the single-person approach since the evaluation will be made based on one person at a time. A viable form of development for this thesis would be using a multi-person approach simultaneously to the personal trainer and the user. However, our trainer's performance was recorded previously; hence it is only necessary to analyze his movement once, store the results and compare them to the person performing in real-time, making the concept more appropriate for real-time usage. Apart from this, it is also necessary that the single-person estimation algorithm achieve the best results possible when dealing with obstacles such as body parts occlusions while minimizing the effects on the real-time performance of the system. When searching for the best algorithm for this application, it was found comparison research made by the BlazePose research team[32] where it compared the state-of-art technologies in this field using mAP(mean Average Precision) and PCK@0.2 accuracy metrics. For this purpose, based on the results provided by the following figure and table, the BlazePose estimator or Mediapipe Pose is used in this dissertation to analyze the poses of the virtual coach and the user with the advantage of being capable of running at over 30 frames per second on mobile devices, and in more powerful devices, even smoother[2].

Method	Yoga (mAP)/(PCK@0.2)	Dance(mAP)/(PCK@0.2)	HIIT(mAP)/(PCK@0.2)
BlazePose GHUM Heavy	68.1/96.4	73.0/97.2	74.0/97.5
BlazePose GHUM Full	62.6/95.5	67.4/96.3	68.0/95.7
BlazePose GHUM Lite	45.0/90.2	53.6/92.5	53.8/93.5
AlphaPose RestNet 50	63.4/96.0	57.8/95.5	63.4/96.0
Apple Vision	32.8/82.7	36.4/91.4	44.5/88.6

Table 4.1: Quality Evaluation of the State-of-the-Art Human Pose Estimation using the PCK@0.2 and mAP(mean Average Precision) metrics[32].

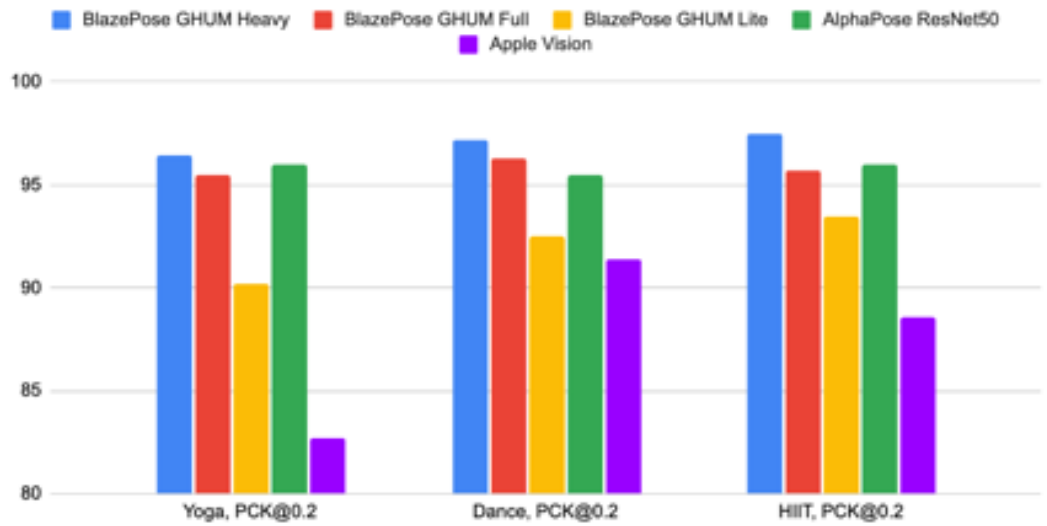


Figure 4.3: Quality Evaluation of the State-of-the-Art Human Pose Estimation using the PCK@0.2 metric[32].

After the comparison with state-of-the-art technology in human pose estimation, it was necessary to assess the algorithm's capability to function in real time compared to its competitors. Knowing this requirement, the BlazePose research team also presented the real-time performance of their algorithms using latency as a metric of evaluation when applied to different types of hardware (e.g., mobile devices or laptops), where the most relevant evaluation was related to the performance of the BlazePose Lite version which has the best performance in terms of latency while maintaining good accuracy.

	Latency	
Method	**Pixel 3 TFLite GPU**	**MacBook Pro (15-inch 2017)**
BlazePose GHUM Heavy	53 ms	38 ms
BlazePose GHUM Full	25 ms	27 ms
BlazePose GHUM Lite	20 ms	25 ms

Table 4.2: The BlazePose algorithm's latency in two different types of hardware.

4.2 BlazePose Detector

With the previous table, it is possible to observe the clear advantage of using this algorithm compared to other state-of-the-art technologies. However, it is essential to understand this particular algorithm's advantage over its competitors. For this purpose, this section addresses how this Machine Learning algorithm works for the reader's understanding.

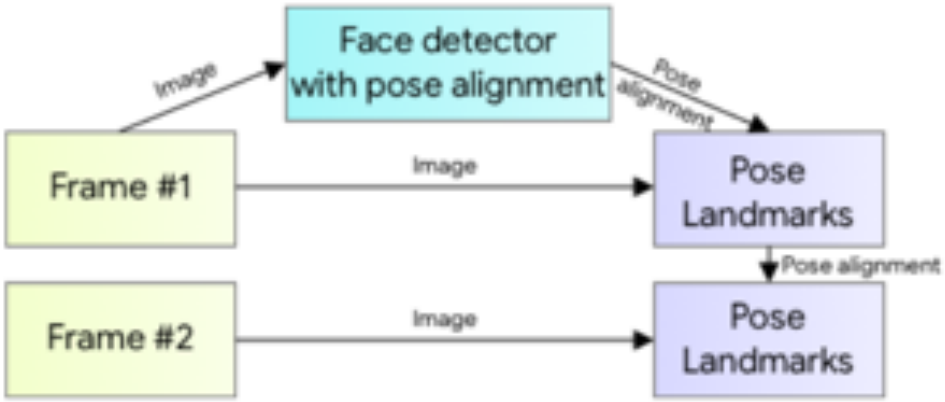


Figure 4.4: BlazePose Detector Inference Pipeline[3].

This improvement over competitors is due to the BlazePose detector-tracker setup using a lightweight body pose detector which in Figure 4.4 is a Face detector with pose alignment, followed by a pose tracker network. The Face Detector is an Machine Learning algorithm tailored for mobile GPU inference. It can run at a speed of 200-1000+ FPS on flagship devices[3]. Apart from this performance, it provides an accurate facial region of interest, 2D/3D facial keypoint or geometric estimation, facial features or expression classification, and face region segmentation. The Detector uses the person's face to obtain the torso's position and additional alignment parameters, which improves the tracker's performance:

- Middle Point Between Person's Hips
- Size of Circle Circumscribing Person
- Incline (Angle between lines connecting the two mid-shoulder and mid-hip points)

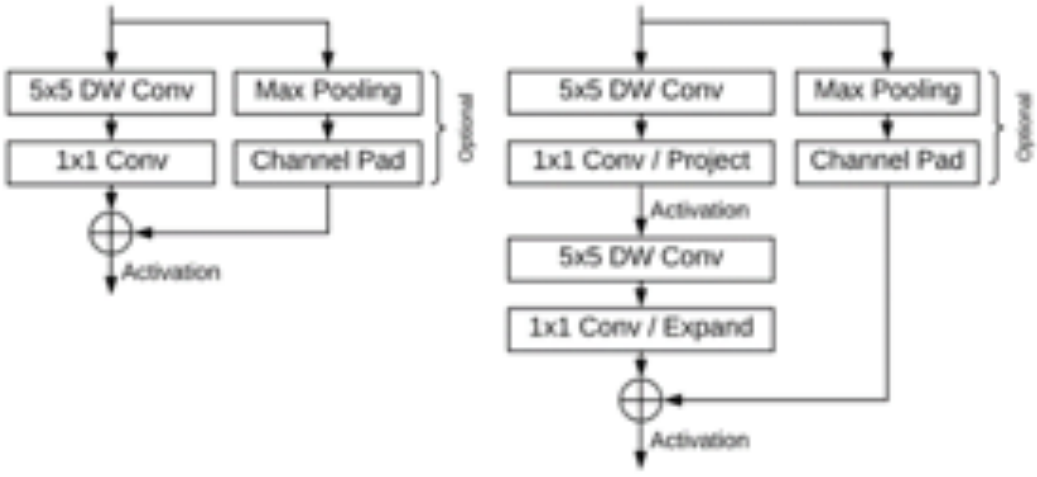


Figure 4.5: On the left side is the BlazeBlock and on the right is the Double BlazeBlock which compose the Blaze Facial Feature Extraction Network Architecture.

Layer/block	Input size	Conv. kernel sizes
Convolution	128×128×3	5×5×3×24 (stride 2)
Single BlazeBlock	64×64×24	5×5×24×1 1×1×24×24
Single BlazeBlock	64×64×24	5×5×24×1 1×1×24×24
Single BlazeBlock	64×64×24	5×5×24×1 (stride 2) 1×1×24×48
Single BlazeBlock	32×32×48	5×5×48×1 1×1×48×48
Single BlazeBlock	32×32×48	5×5×48×1 1×1×48×48
Double BlazeBlock	32×32×48	5×5×48×1 (stride 2) 1×1×48×24 5×5×24×1 1×1×24×96
Double BlazeBlock	16×16×96	5×5×96×1 1×1×96×24 5×5×24×1 1×1×24×96
Double BlazeBlock	16×16×96	5×5×96×1 1×1×96×24 5×5×24×1 1×1×24×96
Double BlazeBlock	16×16×96	5×5×96×1 (stride 2) 1×1×96×24 5×5×24×1 1×1×24×96
Double BlazeBlock	8×8×96	5×5×96×1 1×1×96×24 5×5×24×1 1×1×24×96
Double BlazeBlock	8×8×96	5×5×96×1 1×1×96×24 5×5×24×1 1×1×24×96

Figure 4.6: On the left side is the BlazeBlock and on the right is the Double BlazeBlock which compose the Blaze Facial Feature Extraction Network Architecture.

After the detection of a person's presence in the current frame using BlazeFace Detector, the Tracker can predict: (i) Keypoint coordinates, (ii) Person's Presence and (iii) Refined Region of interest on Current Frame. If the tracker indicates no human presence, the detector network will be re-run on the next frame[2]. The tracker or person detector also differentiates from its peers due to a different post-processing step. The majority of modern object detection solutions rely on the Non-Maximum Suppression (NMS) algorithm for the last post-processing step[2]. However, this processing step when dealing with sce-

narios that involve highly articulated poses like humans breaks down due to the fact of multiple ambiguous boxes that satisfy the intersection over union (IoU) threshold for the NMS algorithm[2]. The BlazePose Detector overcomes this limitation by detecting the bounding box of a relatively human rigid body part, such as the face or torso being the strongest signal to the neural network of the person's face when describing the position of the torso[2].

When dealing with a person's constant presence, the algorithm saves a significant amount of resources by only performing the tracking task each frame, being amongst the competitors, one of the best choices for real-time usage. Apart from the procedure itself, it is crucial when using a human pose estimator to learn the notation of the different key points or landmarks that constitute the stick figure of the person. The authors have also provided this notation where the notation relies entirely on the skeleton's joints in 3D normalized coordinates where the z coordinate or landmark depth is estimated based on the midpoint of the hips being the origin. The estimator also returns an output related to the visibility of each landmark, which helps identify occlusion situations and, in this manner, prevents them.

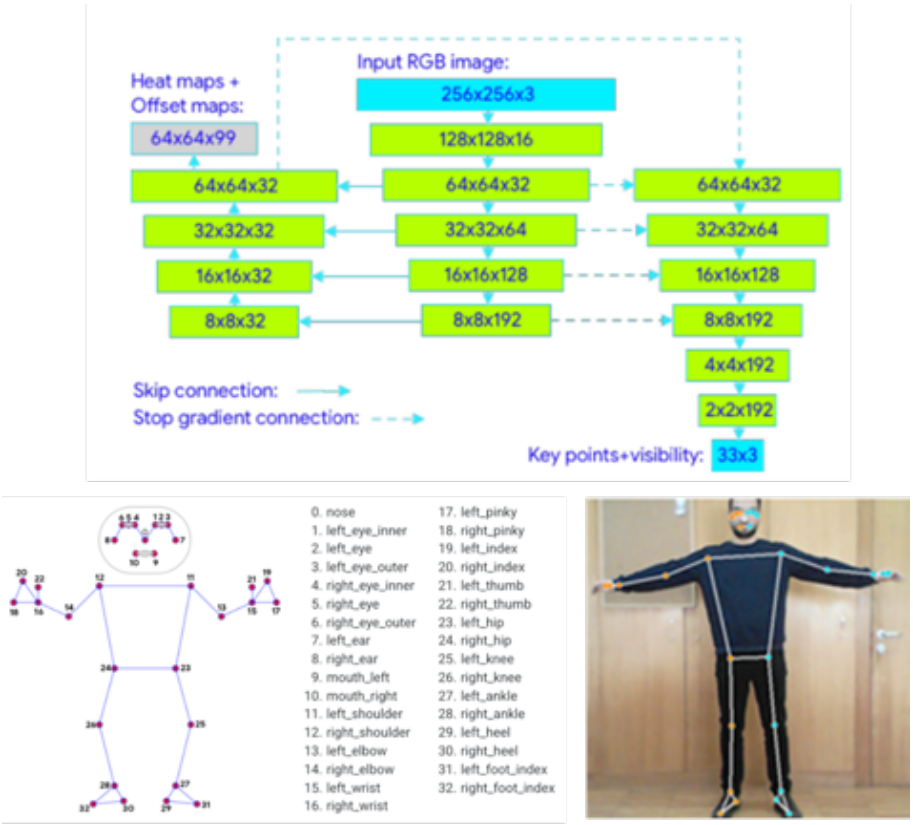


Figure 4.7: On the left side of the figure is the neural network's architecture that predicts the position of the landmarks that constitute the skeleton, and on the right is the output from the BlazePose with the respective notation.

5

Development

With the main objectives and the review of the fundamental concepts, it is clear that the development of this work encompasses several diverse topics to form the system. These topics range from creating an immersive and interactive system and assessment of the user's condition after careful consideration by a medical professional to the promotion of physical exercise while providing a component of entertainment. This chapter explains the development of the interactive questionnaires, personal trainer exercises, exercise analysis, relevant data for the system, the pose comparison between the user and the personal trainer and the adaptation mechanisms of the system.

5.1 Interactive Questionnaires

One of the critical components of the evaluation of sarcopenia is the questionnaires that supply information regarding the user's current condition. As mentioned, there are two main evaluation questionnaires:

- SARC-F: Presence of sarcopenia
- Rikli-Jones: Category of severity

One of this dissertation's main goals is to develop an interactive and immersive system that can provide knowledge and entertainment. Questionnaires can be very tedious and time-consuming, damaging the interaction with the user. With this concern, the implementation of these questionnaires is in a manner that will require interaction with the user in real-time by displaying the necessary information regarding questions and possible options selected based on the user, more specifically, the hands. The information appears on the screen with the support of voice lines that will read the questions in case the user can not see the questions at hand, which is quite common when dealing with a population like older people. These voice lines are helpful to read the questions out loud and to guide the user when navigating in the system, substantially improving the interaction between the user and the personal trainer.

Stage	Voice Line
SARC-F Inicial Menu	I can see that you haven't perform the SARC-F Questionnaire, let's do it.
Rikli-Jones Inicial Menu	I have detected the presence of sarcopenia in you, we need to perform this questionnaire as well.

Table 5.1: Table with the complementary voice lines that provide instructions to the user while performing the questionnaires.

The voice lines rely on neural voice technology from IBM. This type of technology synthesizes human-quality speech from input text using a variety of deep learning networks. The product then uses three Deep Neural Networks to predict the acoustic features of the speech to encode the resulting audio[23]. For each of the phones in a sequence of phones to be synthesized, the acoustic model uses a decision tree model that considers the phone in the context of the preceding and following two phones. Using these, it can produce a set of acoustic units, reducing the complexity of the search since it only relies on units that meet criteria of contextuality.

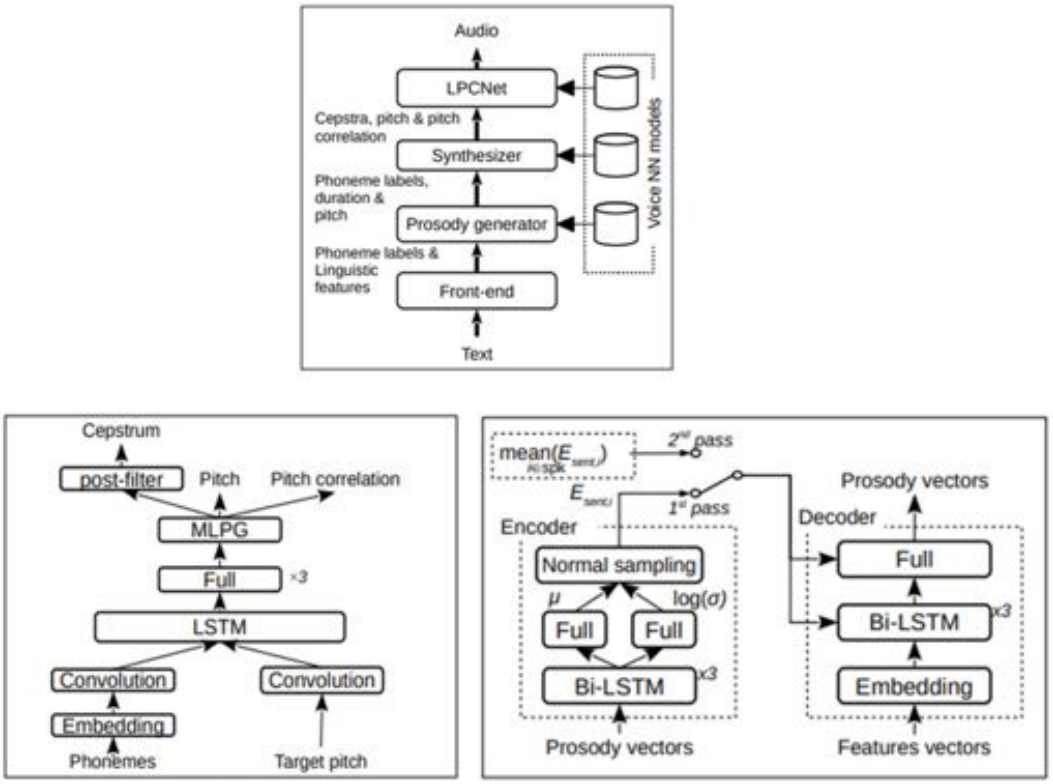


Figure 5.1: The Watson text-to-speech from the IBM implementation pipeline, where the top row image contains the general pipeline from the input text to the output audio. The bottom row contains on the left side the Synthesizer network, and on the right, the Prosody generator [23].

The detection of the hands was possible by using cvzone[1] which incorporates the MediaPipe hand detector that tracks the hand keypoints of the user[54]. This hand detector relies on a pipeline that uses two models, a palm detector that supplies the bounding box that contains the hand and a hand landmark model that uses the cropped hand bounding box to predict the hand skeleton[54]. With the cvzone library, it is also possible to define the key points that are important for the application, which in this case are the index fingers of both hands. With the index fingers, it is possible to define a collision function that will interact with the available options for the questionnaire, allowing the person to respond quickly to the questions. Once the mechanisms of interaction are working, the person can fill out the questionnaire where in the end, the estimation of the score based on the answers will provide a final verdict that reveals if sarcopenia is present in the case of the SARC-F and the severity of the condition with the Rikli-Jones questionnaire. Storing the person's final scores in a file at the end provides a practical use for future implementations(e.g. using the data of creation of the file, it is possible to create a weelky periodic function that reassesses the person's condition) and allows the system to redirect the user to exercises that are the best fit for their condition. With the stored values, the user only performs this task once, allowing the user to focus on more important tasks inside of the system.

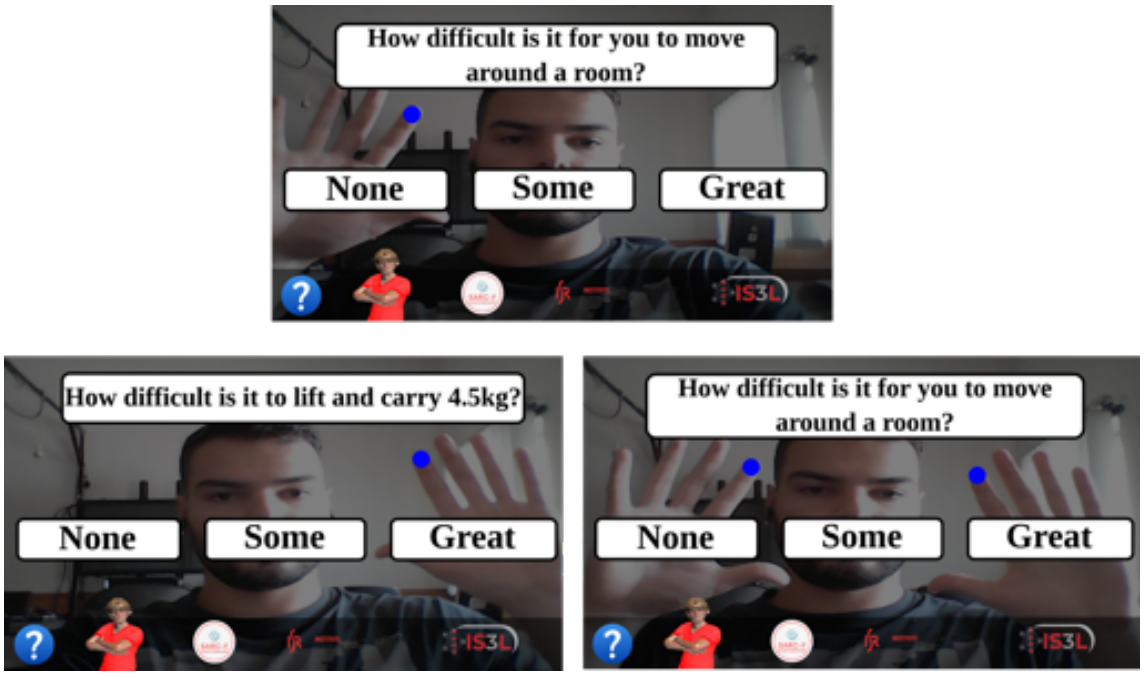


Figure 5.2: Real-time Interaction Mechanisms using the cvzone library, specifically the hand detector. These mechanisms work for single-hand usage (top row) and, if necessary, use both hands simultaneously (bottom row).

[1]Computer vision package to run the Image processing and AI functions using OpenCV and MediaPipe libraries.

5.2 Personal Trainer Exercises

Once the system can recognize the user's presence and the severity of the condition, the personal trainer needs to provide the exercises that suit the user's needs. After careful research on the topic, it is possible to define a good amount of physical activities per category that target different muscle groups, supplying a diverse catalogue of exercises. With the movements relevant to this work, it is possible to use the optical motion capture system to record a certified actor while performing these, providing the correct performance of the movements and therefore supplying a reliable animation that can be transferred to the avatar that can serve as an example performance. To accomplish this objective, the technician must follow a pipeline of procedures to provide the best possible outcome from the optical motion capture, transfer of the motion to the avatar, and gather relevant information based on human motion analysis.

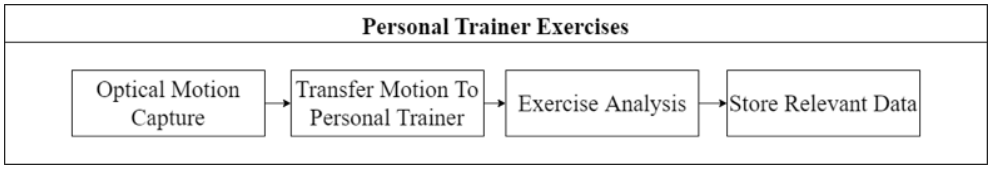


Figure 5.3: Implementation Pipeline of the personal trainer exercises.

5.2.1 Optical Motion Capture Procedure

As mentioned previously, optical motion capture received increased interest over the years, making it a standard in industries requiring animations with natural human characteristics. However, providing this level of detail requires an extensive and specialized creation pipeline, going from the hardware setup to the final product delivered by the software used for the animation. Seeing that Institute of Systems and Robotics provided the hardware setup and physical equipment necessary to record the actor, in this document, the main focus will be on describing the pipeline from the available hardware and good practices to the file delivered by the software to incorporate into the avatar. The following flow chart shows the main steps that need to be accomplished by order of performance to obtain a quality animation.

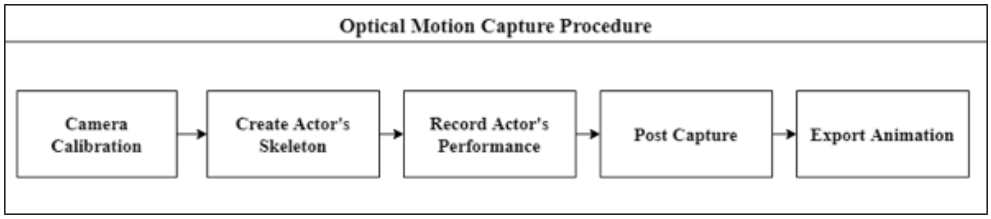


Figure 5.4: Flow chart with the order of performance from left to the right side, exemplifying the key steps of Optical Motion Capture using passive markers.

Although this flow chart explains the main steps to capture movement, there is an extensive list of requirements and good practices that a technician and actor need to comply with to ensure the best tracking results that the system can provide. Starting with the good practices that a technician needs to follow, the following topics are related to the preparation of the capturing room area to prevent the presence of artefacts, performed before the actor enters the area:

- Removing obstacles that obstruct the camera's views.
- Minimizing the system exposure to light interferences such as sunlight.
- Masking reflective objects in the capture room using physical procedures.

After careful preparation of these capture area procedures, it is necessary to perform the camera calibration phase so the system can accurately detect and track the MoCap marker's suit. This procedure is performed on the capturing software using methods that rely on the usage of specialized instruments(e.g., calibration wand and triangle) designed for this purpose, where calibration can occur using the following steps:

- Remove objects that physically interfere with IR light by masking procedures embedded in the software.
- Wanding calibration wand with markers in a specific configuration for sample collection.
- Definition of the Ground Plane to refine camera position and orientation.

The wanding process aids in camera calibration by utilizing a specific markers configuration present in the wand to reflect the several IR-rays emitted by the cameras, pinpointing their locations. It is crucial to change the position and orientation of the wand as it is performing calibration so the software has an idea of the capture room layout. The system, while recording, will provide feedback on metrics related to the calibration quality so the technician can perform it correctly. After camera calibration, it is necessary to define the ground plane of the capture volume by utilizing a proprietary calibration triangle where the vertices are composed of markers. Orientation of this tool is vital since it will provide the information related to the X, Y and Z axes, where the long side of the triangle is the positive Z axis, and the shorter side, the positive X axis using the Y axis pointed upwards. After the precise definition of the camera's positions and the correct estimation of the ground plane, the actor can wear the MoCap suit that contains the passive markers in a specific configuration.

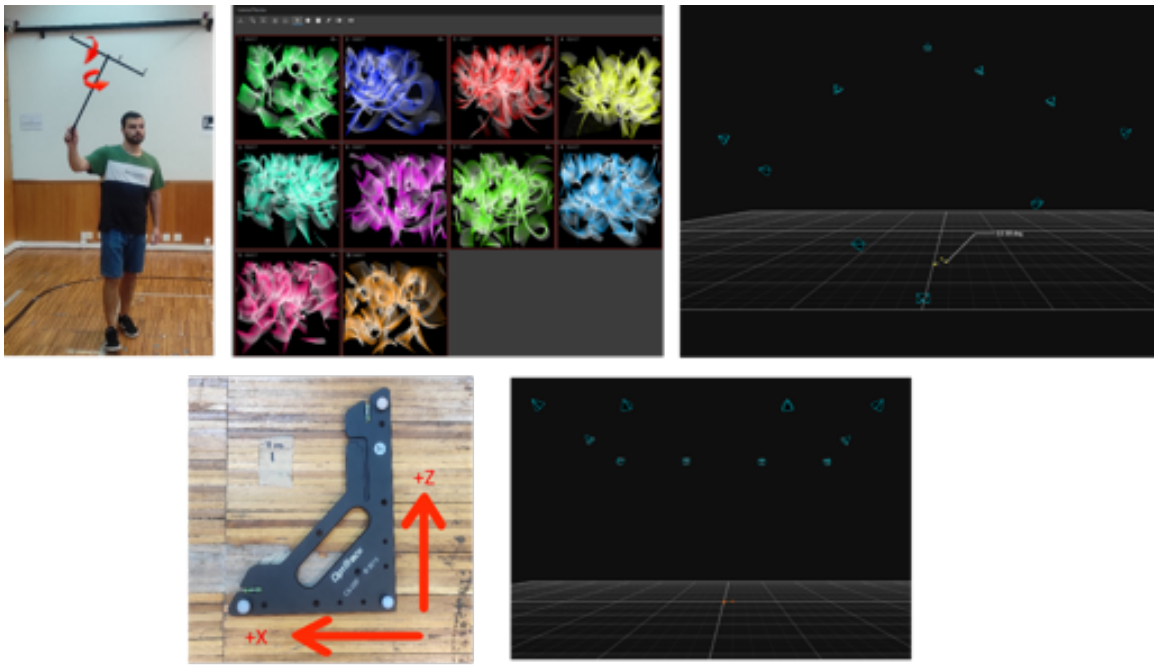


Figure 5.5: The top row presents the wanding process, and the bottom row shows the calibration triangle with the respective axis to define the ground plane, aligning the cameras according to it.

On the Optitrack software, Motive, there are several choices in terms of marker sets to be selected that differ in the number of markers and positions in the suit itself, allowing the system to be versatile in their usage(e.g., record only the upper body or obtain a more precise estimation using more markers). In this dissertation, the baseline marker set using 37 markers was found to be sufficient since it provides the level of detail required. However in future works, if the animation needs to be more representative, there are other marker sets in the software that use more markers, for example, the biomechanics marker set Rizzoli Body with 43 markers.

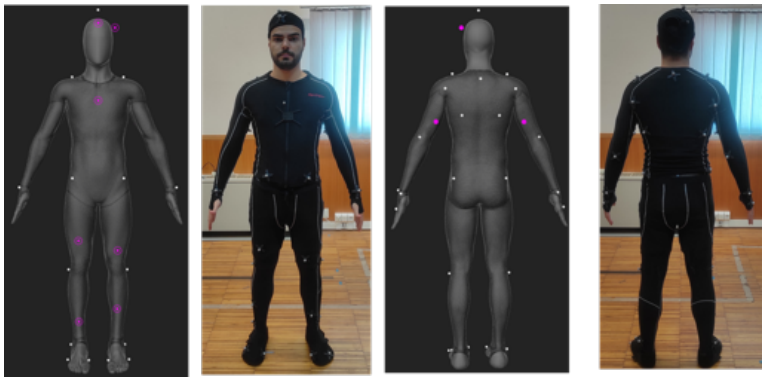


Figure 5.6: Configuration of the marker set to record the animations incorporated into the avatar.

Following the recording session, depending on the type of movement, unstable estimation, occlusion, or unlabeled markers will be a common occurrence when a system has insufficient cameras or markers. In animating articulated figures, we are often concerned with signals defining joint angles or positions of joints[5], which is the case of this system where the markers' positions over time are signals containing values of each frame of the capture session. With this assumption, dealing with the previous issues can be seen as a signal processing problem where markers' positions, along time, will define a spline curve with gaps or jitter due to missing data or inaccurate estimation. Software embedded methods manage these occurrences with interpolation functions(e.g., linear, cubic) that use the points before and after the gap and smoothing functions for jitter based on low-pass filters with customized cut-off frequency.

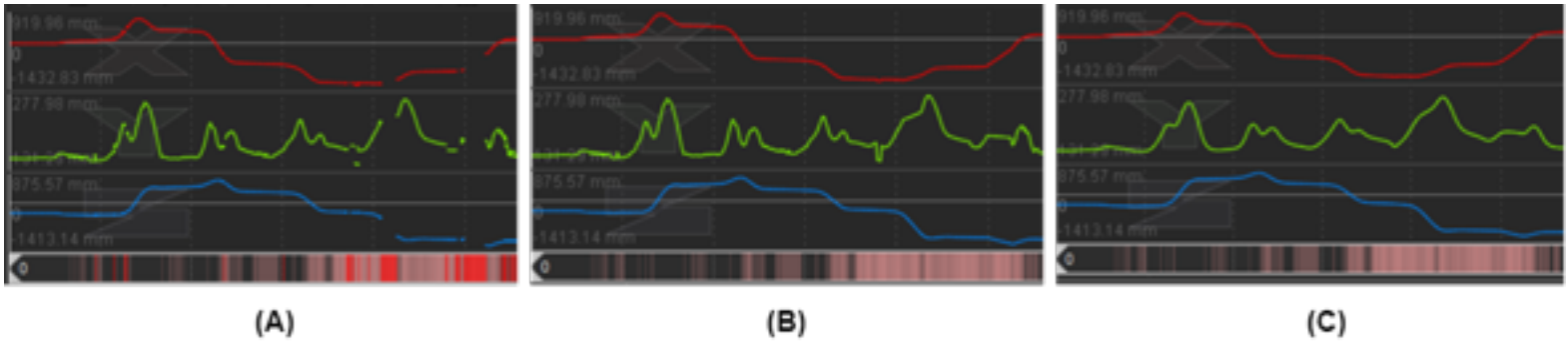


Figure 5.7: Example of the post-capture method used to improve the overall definition of a curve related to a joint using linear interpolation(A->B) and a low-pass filter with a cut-off frequency of 3 Hz (B->C).

Once the capture and post-capture steps are complete, the software allows the animator to export in several formats that differ in terms of the information provided. From the table below, it is possible to observe the formats with the most information about the capture session are the FBX[2] and CSV (Comma-separated values) format. Both provide reconstructed 3D marker data and, most importantly, the skeleton data established by the markers that will form a connection between the performance of the actor in the capture session and the pre-defined skeleton of the avatar. The table also shows another helpful format, BVH[3], which only supplies skeleton data being ideal in terms of optimized usage. However, due to incompatibilities between software when using BVH or CSV, FBX was chosen as the primary format of the implementation, delivering the skeleton data and afterwards merging it with the virtual personal trainer skeleton.

[2]FBX (Filmbox), a proprietary file format developed by Kaydara to supply 3D geometry and animation data represented in ASCII or Binary data.

[3]BVH (Biovision Hierarchy) originally developed by Biovision, a motion capture services company to provide motion capture data.

Tracking Data Type	CSV	C3D	FBX	BVH	TRC
Reconstructed 3D Marker Data	x	x	x		x
6 Degree of Freedom Rigid Body Data	x		x		
Skeleton Data	x		x	x	

Table 5.2: Table with export data types available in the software, showing the information provided by each of them[37].

5.2.2 Personal Trainer Motion

After the recording session in the motion capture room, the FBX file that contains the motion data needs to be transferred to a 3D humanoid model. The Mixamo[4] humanoid character facilitated the process of animation by providing a mesh and a gold standard humanoid rig, avoiding in this manner the definition of these which can be a very detailed and time-consuming process. Although one problem remained, the motion data could not be transferred directly to the character since the motion. Optitrack provides a streaming mode where the motion data can be streamed onto external applications(e.g., Motion Builder, 3ds Max) in real-time with the assistance of streaming plugins[37]. After exploration of this option, it was clear that the motion data needed post-processing due to jitter and other issues related to the MoCap markers occlusions resulting in unrealistic animations. After motion data post-processing and exportation of the file in FBX format, one of the most viable options to perform this data transfer is using secondary software, MotionBuilder[5]

Using MotionBuilder, it was possible to define rigs that complied with standard criteria regarding rigging and more. Seeing that the types of exercises required different environments settings and conditions, this software allowed for the manipulation of joints, bones, and objects that were not possible to define only using Motion Capture(e.g., the definition of hands movements due to lack of markers), providing a level of detail which was not possible without it. Apart from these advantages, the software can modify animation using the rigs with forwarding, inverse kinematics, and keyframing techniques. These mechanisms are beneficial as the movement is no longer defined only by the capture session, allowing modification based on the animator's needs or interactions with objects such as weights or chairs.

[4]Mixamo is a computer graphics technology company developing services for 3D character animation, supplying animation sequences and several 3D character models that differ due to morphologies.

[5]Professional 3D character animation software by Autodesk primary purpose of virtual production, motion capture, and traditional keyframe animation.

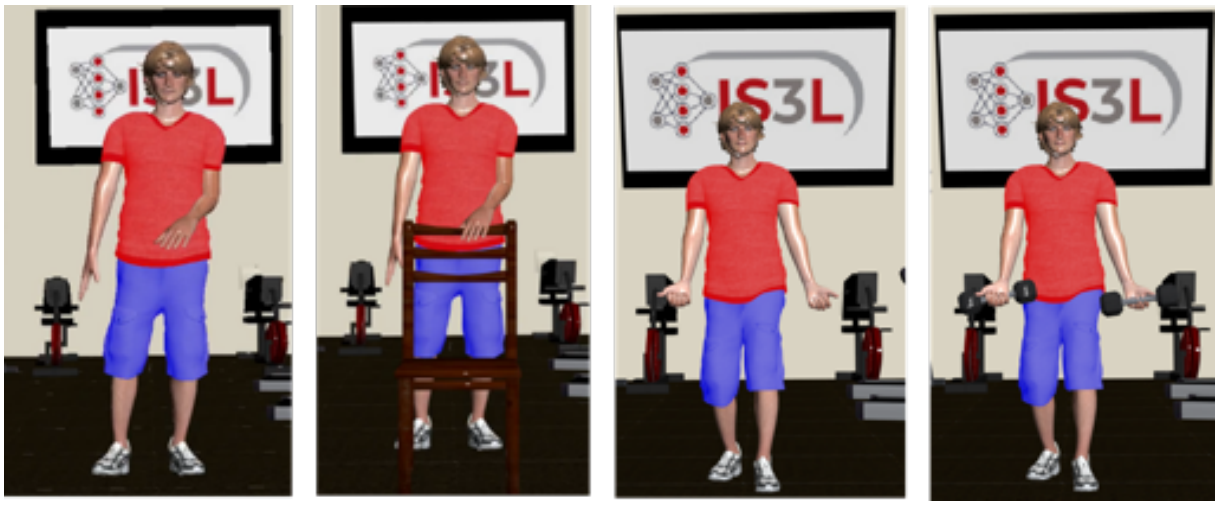


Figure 5.8: Two animations with and without objects where it is possible to observe the difference in terms of realism.

5.2.3 Exercise Analysis

After the detailed task of animating the virtual character, the evaluation of the personal trainer and user performance needs to be in a form that generalizes the exercise using the least possible amount of information. For this purpose, there are several ways to accomplish this analysis; however, in this dissertation, only the three main approaches when developing this work will be discussed, showing all the advantages that the final evaluation techniques provide. The main approaches are:

- Euclidean Distances
- Adaptive 2D Grid
- Static 2D Grid

Before the main approaches, it is crucial to discuss the reduction of the amount of information processed by the system in real-time. One way to accomplish this is by reducing the number of variables being monitored in real-time when observing the motion. As mentioned before, there are several outputs for each landmark that the pose estimator supplies(e.g., 3D position, visibility); however, after experimenting with the estimator for several exercises, it was possible to observe that the description of the movements in most cases is sufficient using only two dimensions without losing essential information. Nevertheless, to accomplish this simplified analysis, first of all it is necessary to convert the normalized coordinates in three dimensions into image coordinates in two dimensions.

5.2.3.1 Application of the Estimator on 2D Image Plane

BlazePose is a human pose estimator that provides four outputs:

- Landmark coordinates (x, y) in the image plane normalized between 0 and 1 accordingly to the image height and width.
- Landmark depth z calculated assuming the depth at the midpoint of the hips as the origin (the closer the value of the depth, the smaller it is).
- Visibility is a value with a range of [0.0, 1.0] that indicates the likelihood of being visible in the image.

The landmark's depth uses the depth at the midpoint of the hips as the origin, becoming the value smaller as it comes closer to the camera[2]. Based on the output provided by the human pose estimator, a viable way to create the evaluation algorithm would be using the landmark coordinates (x,y) in conjunction with the landmark's depth z. In this approach, it is possible to observe movements that vary according to the depth axis, for example, if the hand is closer to the body or the camera. It can be helpful in some exercises; however, most of them could be evaluated based on the variation in the x and y-axis related to other body parts, simplifying the whole assessment process and decreasing the amount of data used in the analysis. The neural network gives normalized 3D landmark coordinates, but for the objective of working on the image plane, the system needs the coordinates related to the image dimensions. Although the normalization is valuable and efficient when storing the values or calculating metrics such as euclidean error, the coordinates of the key points based on the image are more discriminant allowing the assessment of the person's performance. Undoing the normalization of these outputs computes this information using a formula that considers the image's dimensions.

$$value = value_{norm} * (value_{max} - value_{min}) + value_{min}$$

5.2.3.2 Euclidean Distances

Pose comparison between poses has seen several approaches with different methodologies, using Euclidean distance, the difference between angles of the joints, or Discrete Fourier Transform (DFT) that identifies periodicities in the data to compare them, used mainly in action recognition. Euclidean distance was one of the implemented approaches since it is the simplest one while fulfilling the evaluation purpose. This method was developed by calculating the distance between different body parts to execute the action. For example, the Euclidean distance between the hands and the shoulders was significant when executing a bicep curl since this varies significantly over time. With these types of distances and an algorithm based on thresholds, it would, in some exercises, be sufficient to evaluate the execution.

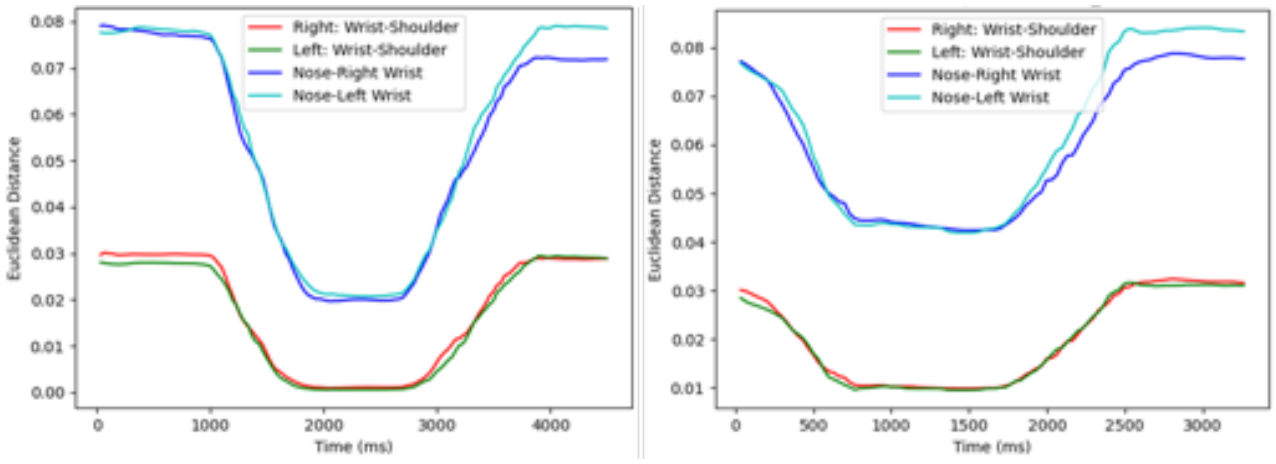


Figure 5.9: Euclidean distance method exemplifies the difference between a good(left side) and bad(right side) Bicep Curl execution.

Although this hypothesis works for several different exercises, some issues originated from following this method. One of the significant problems observed was the disparity of euclidean distance values between people of various propositions. Thresholds that evaluate the exercise depend on these values, originating the need to adjust the thresholds for a portion of the population but worsening the system's overall performance for others. Although the thresholds can describe some exercises, there is no way to assess whether this occur do the correct positions of landmarks or if it happens in a different pose than what was asked for by the user. For example, suppose the position indicated by the personal trainer is to place the right foot in front of the left foot. Theoretically, the user could put the left foot in front of the right one since there is not a discriminatory factor based on the body parts.

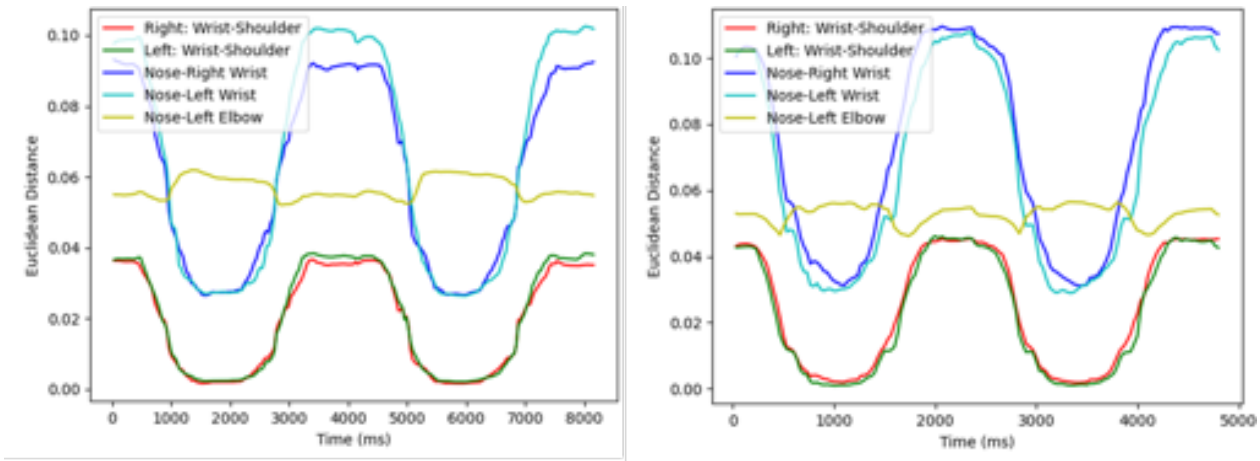


Figure 5.10: These graphs display a Bicep Curl repetition performed by people with different proportions. The left graph shows the smaller person's performance, and on the right side, the larger person's performance.

5.2.3.3 Adaptive 2D Grid

After noticing the disadvantage of the euclidean distance approach, future approaches needed to be discriminatory in terms of body parts used in the performance of an exercise. The second approach developed in this work had that objective, the adaptive grid. This type of grid was a set of lines defined and updated in real-time using the 2D position of anatomical landmarks(e.g., the nose from a middle vertical line or the shoulders for a horizontal line) from the skeleton joints supplied by the BlazePose to create each cell of the grid. With the cells or regions, the system could detect the presence of the landmarks of interest over time. With this information, the system can monitor the trajectory of the landmarks of interest using the regions to describe the performance and, therefore, an exercise buffer. This method provides a simplified exercise description, and in theory, a universal evaluation since the grid is formed based on the person. The personal trainer's performance can act as a baseline for the system to evaluate other individuals using the stored exercise buffer, labelled using the name of the exercise. When trying this hypothesis, the system could create and update the grid in real time without affecting the overall performance. However, the grid had one major disadvantage, recurrent collapsing regions. This problem was due to the creation of lines based on the positions of moving or non-static joints. One primary example of this phenomenon was when evaluating a repetition of a sitting exercise where the hip line would overlap other lower lines of the person when the person reached the midpoint of the repetition.

Figure 5.11: Virtual personal trainer performing the lifting from the chair exercise, originating collapsing regions and therefore affecting the evaluation of the exercise.

5.2.3.4 Static 2D Grid

Although the adaptive 2D grid had some disadvantages, the grid procedure still is a viable form of analysis due to its discriminatory nature. With the regions, it is possible to differentiate orientation(e.g., left/right side) and trajectories along the time of the landmarks of interest while providing a discriminatory method for people with different proportions. Establishing these new discriminatory regions requires the positions of the selected joints and relative distances based on these to form the lines, creating a grid that contains more regions and, therefore, is more accurate for the exercise evaluation. After experimentation, one form of the grid came to mind, a static grid. With this type of grid, the issue of recurrent collapsing regions disappears since they are constant over time. However, to form the grid, the system needs to estimate the most representative values of each joint of interest based on the current person's position, allowing the person to create their exercise zone. After extensive experimentation with selecting the joints of interest for the grid, the conclusion was that the formation of the grid could not be only joints because, in some cases, the change of the landmarks of interest, over time, would be too subtle. Nonetheless, the resolution of this issue is possible by utilizing other metrics specific to the user.

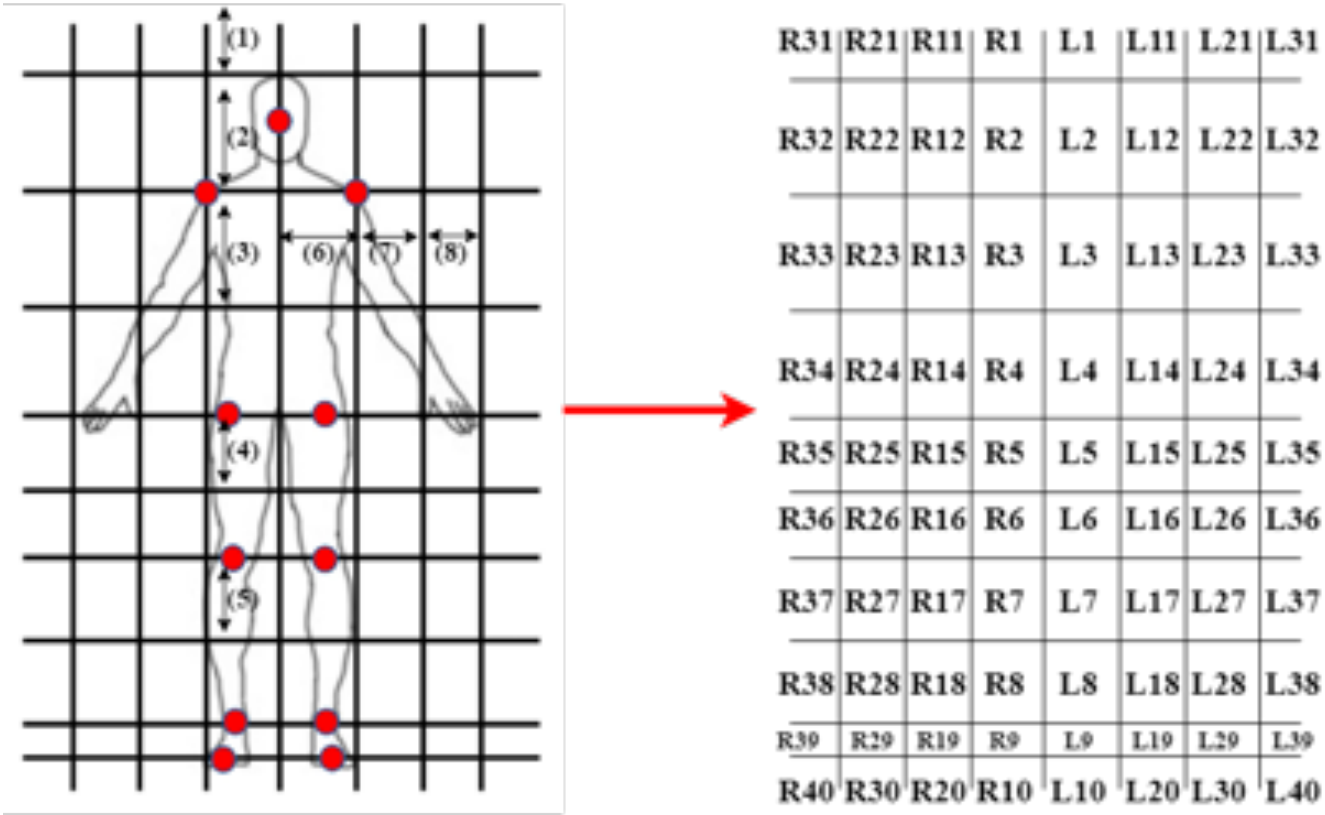


Figure 5.12: Joints selected and relative distances for the grid formation, being on the right side the notation for the regions.

Landmarks of Interest	Index
Nose	0
Shoulders	11, 12
Hips	23, 24
Knees	25, 26
Ankles	27, 28
Front Toes	31, 32

Table 5.3: Landmarks Of Interest to form the 2D grid with the pose landmarks indexes.

Index	Relative Distance
1	$(\text{nose}_y - shoulder_y) \cdot \frac{9}{2}$
2	$(\text{shoulder}_y - nose_y) \cdot \frac{3}{2}$
3	$(\text{shoulder}_y + hip_y) \cdot \frac{1}{2}$
4	$(\text{hip}_y + knee_y) \cdot \frac{1}{2}$
5	$(\text{knee}_y + ankle_y) \cdot \frac{1}{2}$
6	$(\text{rightshoulder}_x + leftshoulder_x) \cdot \frac{1}{2}$
7	$\text{leftshoulder}_x + 1.2 \cdot Index_6$
8	$\text{leftshoulder}_x + 1.2 \cdot Index_6$

Table 5.4: Relative Distances ordered by the indexes in the previous figure.

The gathering of the joint values is different for the personal trainer and the user since it knew beforehand that the coach would not move from his current location. For the personal trainer, the values that form the static regions use the values seen on the first frame of the exercise in video format. While for the user, it is important to assume that the person will move around the exercise zone, resulting in a more detailed calibration process for these values, where the whole process is in the chapter relative to the comparison of poses between the coach and the user. With the grid formation, the system can perform the exercise analysis using a fraction of the information supplied by the pose estimator when running it on a correct repetition of each exercise.

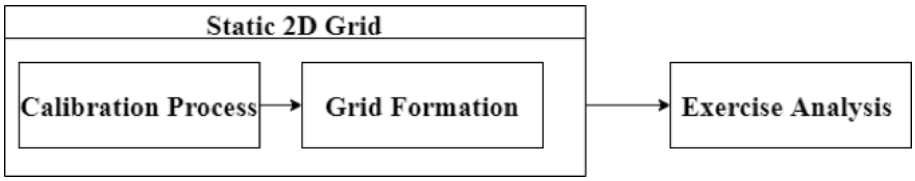


Figure 5.13: Final approach for the formation of the grid which will serve as a guideline for the exercise analysis.

5.2.3.5 Relevant Data

As mentioned before, most exercises have the evaluation from a single constant camera perspective, with a static camera in front of the person or coach. With this observation and with the concern for optimization of the system performance, the exemplification of the exercises will be in video format since it can provide the information necessary for the user and the system while improving the performance when compared with the alternative of placing the personal trainer in a virtual environment while performing the same evaluation from the same camera perspective, saving significant amounts of processing power for other tasks.

The relevant data for a system can vary accordingly to the objective, the equipment that composes a system and the requirements the system must follow. However, a real-time system not only addresses these rules but also needs to consider the amount of memory used while running to avoid common issues like memory shortage. With this concern, when the system's exercise analysis is running on the personal trainer, it only needs three parameters:

- Specification of the landmarks of interest
- The trajectory of the landmarks using the notation of the regions that compose the grid
- Identification of the exercise

Specifying these parameters allows the system to track its trajectory during the performance while associating it with the exercise at hand. At the moment, the technician specifies these parameters manually after carefully observing the trajectory periodicity of the different landmarks that compose the skeleton; however, the definition of these only occurs once, providing the system with the ability to use this information to evaluate the user's performance of the exercises.

Another helpful feature this approach offers is tracking when a landmark of interest is in a region of interest and when a transition between these regions occurs using timestamps. With this information, the system can display or interrupt the video showing the exercise to provide meaningful interaction between the user and the coach. For example, the personal trainer can stop at a critical point of the movement (e.g., in the middle of the repetition) and wait until the person reaches this point to continue the execution of the exemplification. This feature allows the person to follow along with the coach simulating a typical real-life interaction when instructing a person on how to perform a movement. For this purpose, the system needs to extract and store the timestamps when the transitions of regions occur and find a point where it is safe to interrupt the video to ensure that it stops at the correct timestamp. The system to estimate this value uses the timestamps of the transitions when it enters and when it leaves the region to compute the mean timestamp for the region at hand. Using the mean value ensures a safe timestamp for this feature since the equipment used to run the system can vary on computational power affecting the velocity at which the video runs. After the estimation, the system can store these values in files with the identification of the exercises and different directories based on the categories in which they fit.

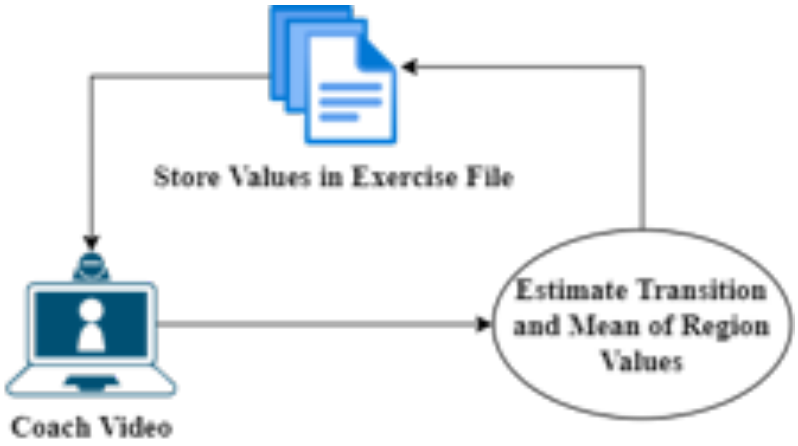


Figure 5.14: Diagram showing the process of timestamp extraction from the video performance of the coach.

5.3 Pose Estimation

With the analysis of the exercises completed exemplification, the system can now analyse the user's performance since the pose estimator is the same and the requirements established for the personal trainer apply to the user, which means that the user, after the calibration process, can perform the exercises assuming that it has the same orientation toward the camera as the avatar. However, as mentioned before, some concerns involve occlusions when using props to perform the exercises. One attempt to avoid this, which can affect the performance and efficacy of the estimator, was the strategic placement of props when needed to perform the exercises in a way that will not occlude any relatively rigid body part, which is essential to estimate the skeleton, like the face or the torso being this an assumption made by the authors of this estimator in the case of single-person use case[2]. The assumption mentioned above is also in mind when evaluating the exercises. Since the system setup consists of single-camera use, the face must be present in the image plane, being necessary that the user is in frontal view or faced 45 degrees away from the camera, improving the algorithm's performance without affecting the execution of the exercise. In the exercises with props that could not be moved, a deteriorating performance from the estimator on the user/avatar can be seen, which will be discussed in the chapter regarding results. The analysis of the user's performance resembles the avatar, meaning it is possible to define a good/bad repetition with the exercise buffer and the landmarks of interest since the user and the avatar perform the movements in a similar environment allowing for the simplification of the overall system and universal behaviour due to the formation of the static 2D grid using relative distances based on the characteristics of the person. Regardless, it is important to explain the whole procedure of the comparison between these poses, which is the main focus of this dissertation.

5.3.1 Comparison Between Poses

As mentioned before, the calibration for the user differs from the avatar's, meaning that the calibration phase will be a time interval where the person needs to perform the T-Pose for ten seconds. In this time interval, the person needs to be immobile and remain in the zone where it will execute the exercise so the system can accurately form the grid. At the same time, the system needs to track minor variations that a person performs during this interval while maintaining the pose. The mean of all the values collected along the time can compute a robust estimate of the joint's position and, therefore, the lines that compose the grid. To avoid the person's movement during this phase, the person needs to place the landmarks of interest(wrists and ankles) in the correct regions for the timer to start and maintain it. While performing the calibration, the person can see through a colour scheme when the landmarks of interest are in the respective areas(e.g., green regions when the landmark is present and red when absent), indication using text when recording the values and the timer value in real-time. After the calibration, the regions will be static for the remainder of the exercise execution, where one point appears in the person's skeleton, representing the person's centre of mass. This point allows tracking not only the centre of mass but also the offset along time that the person has from the calibration due to movement or rotation, activating if the increasing offset passes a certain threshold, the calibration process, allowing the user to establish the new exercise zone without affecting the exercise analysis.

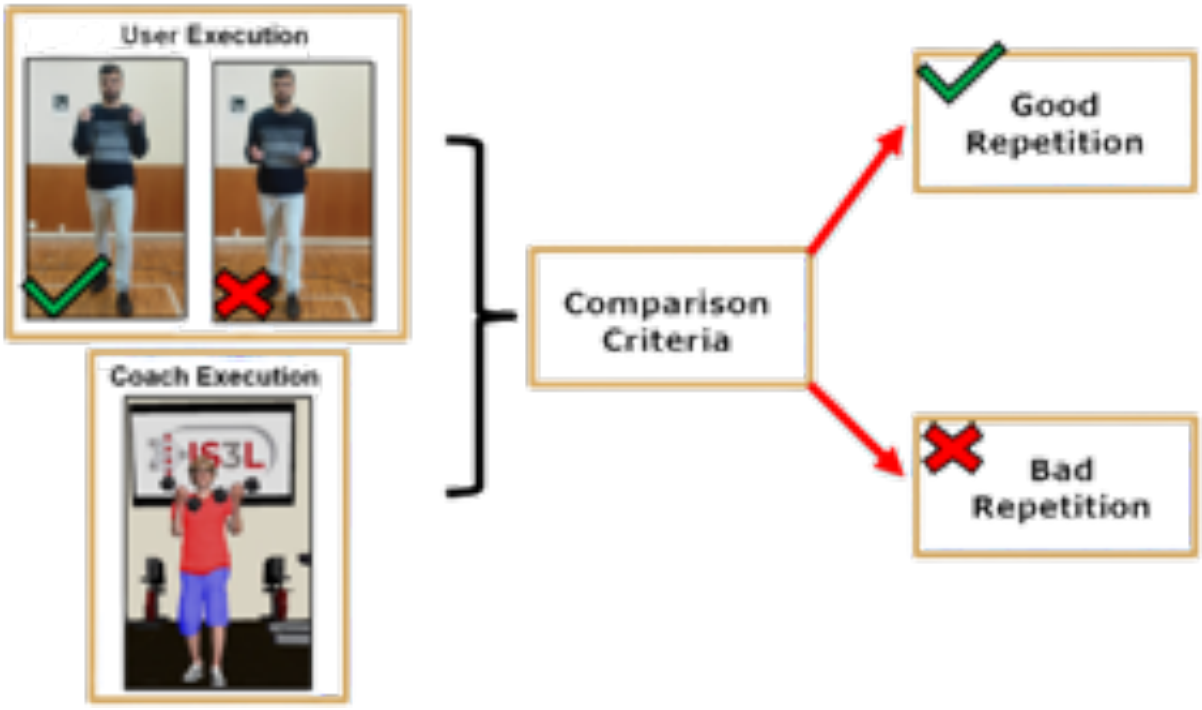


Figure 5.15: Simplified Model of the Comparison Criteria.

Figure 5.16: The figure shows the grid method's appliance to a pre-recorded video of a bicep curl execution on the left side and the right, the grid formation on the virtual personal trainer.

As the user performs the exercise, this descriptor will update in real-time with the regions where the landmark passed and, at the same time, will be compared with the stored exercise buffer provided by the coach allowing in this manner to assess whether the user performed a complete repetition in the correct form. Depending on the exercise, these descriptors can be composed of one or more landmark trajectories since the activity may involve multiple body parts.

The main advantages of this technique are: (i) Simplicity in terms of implementation, (ii) Lower computational cost since it uses a small number of landmarks and a lower dimension of coordinates to characterize an exercise and (iii) Robust analysis of people with different proportions.

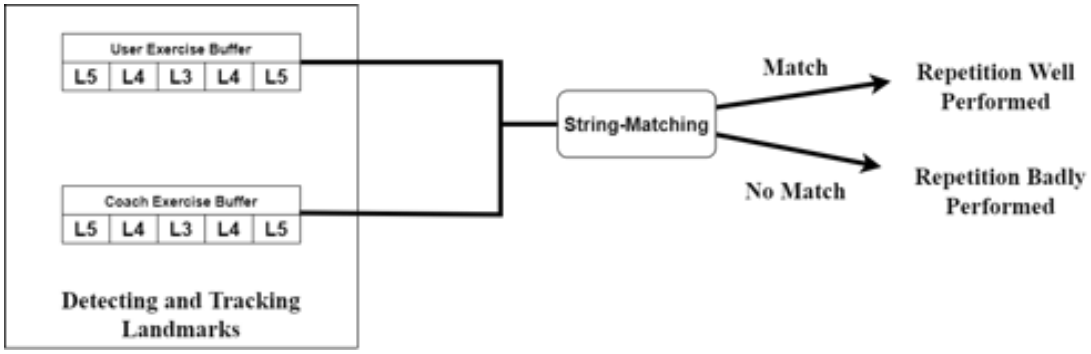


Figure 5.17: Diagram showing the pipeline of the approach implementation where the string-matching comparison method evaluates the person's performance in real-time based on the stored buffer of the personal trainer.

5.3.2 Adaptation of the Exercise

When a person reaches a point of discomfort or frustration by not finishing the exercise, it is crucial to adapt the exercise since the objective is to stress the muscle to a certain degree

to obtain progress but not harm the person to a point where it prevents the usage of the system in the future and this manner affecting the consistency of the exercise routine of the user and overall improvement. To be effective, a coach needs to perform the exercise and adapt the exercise based on the person he is teaching and guiding. Since sarcopenia is a debilitating geriatric condition restricting movement and flexibility, some exercises need to be safer in their execution so the system can benefit the user more than harm. With this concern in mind, several approaches could be followed, but these that will be mentioned are the ones that seem to be the more appropriate: (i) Adapt the exercise to perform it slower and, if used, lower weights and (ii) The coach simultaneous performs the exercise and, if necessary, stops their execution based on the feedback that they receive from the person.

Depending on the severity of this condition, the user can be affected in a specific body part or whole body, being this justifiable on their on to get a medical professional initial assessment to provide this information. Based on this initial assessment, the coach will emphasize the exercises that work on these affected body parts without harming or aggravating the individual by slowing down the exercise and then changing to a different exercise that works on another body part, if not possible. As the coach's performance is in the system as a video format, it is possible to implement a mechanism based on video compression and frame decoding that allows the personal trainer to change the speed of the exercise without a high cost to the performance of the real-time system providing another advantage of using this type of format when showing the exercises. Once the video is playing on a lower frame rate, the coach can provide a voice line related to the weights by informing the user to lower the weight used while performing the exercise, creating a meaningful interaction by giving the user the feeling of being supervised in a customized form.

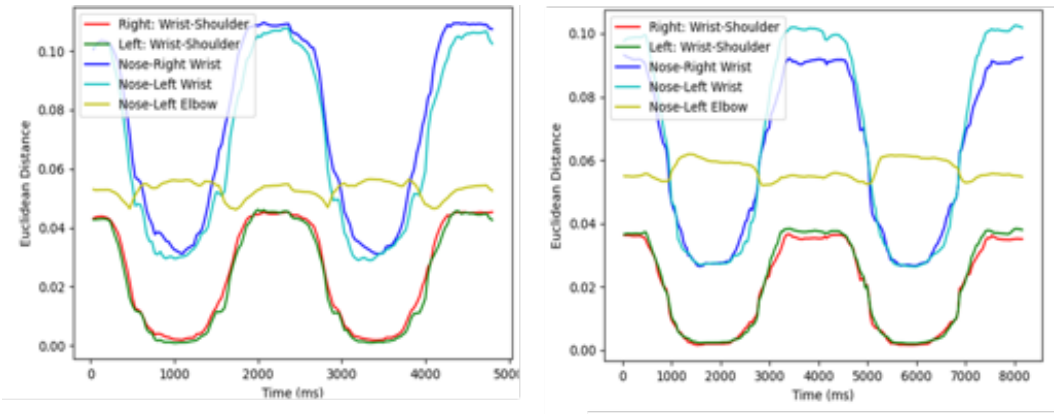


Figure 5.18: On the left side of the figure is the coach's typical performance and on the right is the adapted form of execution for the same exercise. The Euclidean distance metric between the landmark of interest and the body parts of interest, shows it reaching its peak in different timestamps.

As for the third approach, it is possible to perform this using the timestamps, where the system detects a transition between regions by tracking the landmarks of interest and estimating the meantime the coach takes between each region of interest. With the size of the exercise buffer, it is possible to estimate a critical midpoint of execution of the exercise and therefore stop the video of the coach when the system detects that the person is not performing at the same speed as the personal trainer is. When the individual reaches the critical midpoint of the exercise, the video can continue to play until it is completed and, if necessary, repeat the video until the user is satisfied with their performance.

Figure 5.19: Sequence of Events when performing a Seated Opening Arms Repetition, where the coach is waiting until the user reaches the critical points of the exercise.

5.3.3 Application

It is important to mention that one of the main objectives of this work is the development of a system that not only aims for the evaluation of targeted physical activity but also to interact in a meaningful form with the user. With this concern and after the testing phase of the work, an actual demonstration of the overall system allows us to evaluate the system performance with personnel not involved in the technician part of the application to provide insight into the system's acceptability. The implementation of the application relies heavily on the coding language python and the pygame[6] library facilitating some aspects of the overall implementation. Furthermore, the composition of the application depends on the following phases: acquisition of data from the questionnaires, user's selection of the exercise accordingly to the category of the severeness, and monitoring of the user's performance.

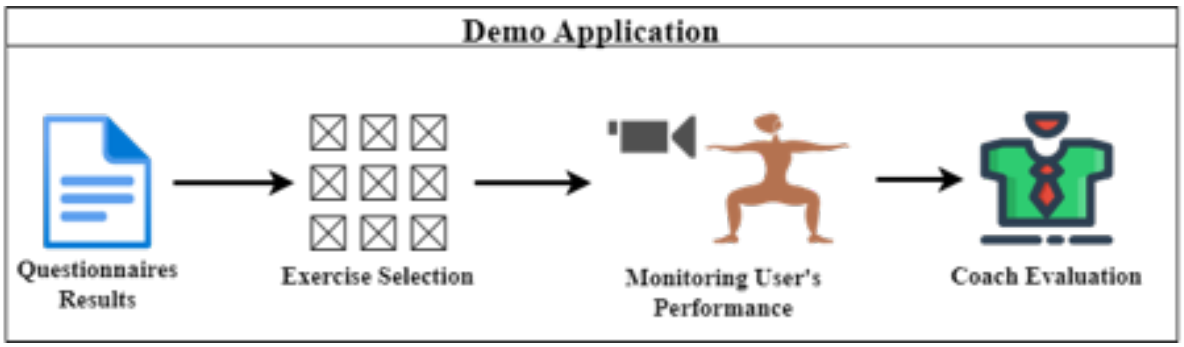


Figure 5.20: Summary of the application's pipeline, highlighting the main phases.

With the previous application pipeline in figure 4.1, it is possible to observe that the first main phase is the search for the results of the questionnaires. Questionnaires are a time-consuming and, many times, tedious task for people to perform where the main objective, in this case, is for the system to determine which types of exercises a person can perform safely. With this concern, the system only needs to perform this step once, saving the relevant data into a file and therefore allowing the person to be more focused on other aspects of the application that are more relevant to the overall experience of the user. After the screening test and evaluation of the severity of the condition, the system can supply the available exercises on the system that the user can select and perform without harming or aggravating their condition. Using the hand detector, the user can navigate through the application using their left index finger and, most importantly, select an existent exercise in their respective category. After the selection, the system shows the coach's performance so the person can perform the calibration procedure and afterwards execute the exercise. While executing the exercise, the system monitors in real-time the user using an RGB

[6]Cross-platform library with the primary purpose of game development

camera, the BlazePose pose estimator, and the evaluation procedure developed in this dissertation. The monitorization allows the system to provide information specifically (e.g., identification of the exercise/category and a repetition counter) to the user's execution.

While monitoring the user, the system provides specific information related to the exercise(e.g., identification of the exercise/category and a repetition counter) and real-time feedback on the user's performance. The system's feedback allows the user to track and improve their performance, being displayed using three mechanisms. A visual scheme that shows the user if he/she is performing the exercise correctly (green colour) or incorrectly (red colour), an audio queue that identifies the completion of a correct repetition or multiple ones. Finally and most importantly, on the bottom part of the screen, the user can see their progress when performing the exercise, presenting the exercise string(grey region notations) and their exercise buffer(green region notations) to show their progress. This feature provides cognitive stimuli since the person can see the region where they need to be based on their current region while displaying the grid regions that do not belong to the exercise string.

In case of prolonged failure, the personal trainer will adapt the presentation of the coach's performance in future executions by slowing down the performance. Furthermore, if the individual cannot perform the exercise in this manner, the system will present one final form of adaptation where the coach accompanies the exercise with the person by stopping the performance at critical points, waiting to continue the performance until the person reaches the point where the coach is. Once the user finishes the execution of the exercise in their respective category, the user can select other exercises, allowing the performance of multiple exercises in one session, working on multiple muscle groups or a specific muscle if it will not harm the user in any capacity.

6

Results and Discussion

Once the system's development is complete, it is necessary to evaluate the human pose estimator performance and the exercise evaluation method when the coach performs the different exercises and the system's capability to provide an immersive and interactive experience to the user. For this purpose, this chapter aims to present the performance of the exercise evaluation and the pilot study results carried out on ISR.

6.1 Exercise Evaluation Performance

When evaluating the exercise evaluation performance, it needs to consider the cases where it can not accurately monitor or define concrete trajectories of the landmarks of interest in real-time. This phenomenon can be too numerous reasons(e.g., the exercise performance at high speeds); however, one of the main reasons mentioned in previous chapters is occlusions. Occlusions in pose estimation algorithms are the main detriment to the joint estimation accuracy, affecting the exercise evaluation. Another factor that can increase the cases of failure of the exercise evaluation is too subtle changes when performing the exercise, not allowing the exercise buffer to be updated since the landmark of interest never leaves the initial region. For these reasons, the leading causes of failure for this evaluation procedure are: (i) Unpredictable Trajectories of the Landmarks of Interest, (ii) Obstacles Occluding Person's Body Parts and (iii) Subtle Changes When Performing the Exercise.

From 25 different exercises in the system, the (i) and (ii) causes of failure affect three exercises. Unpredictable trajectories do not allow the evaluation of the Plantar Flexors exercise from the Strength exercises and the Hip Extension exercise from the Elevated/Moderate Functionality exercises. While Subtle changes only affect one exercise, Raise Heels And Toes, from the Low Functionality exercises since the heels or toes never leave their initial region. On the other hand, occlusions affect the same amount of exercises as the other two reasons, affecting three exercises due to occlusions of body parts during exercise execution. Leg Extension from the Strength Exercises, Leg Flexion from the Low Functionality

Exercises and Dorsiflexor from the Strength Exercises are affected by this phenomenon as these exercises require a chair in front of the user. From this information, it is possible to see that the developed comparison criteria evaluate 92% of the exercises when not considering obstacles occlusion and 76% when including obstacles occlusions. Examples of the cases of failure are in the appendix of this dissertation.

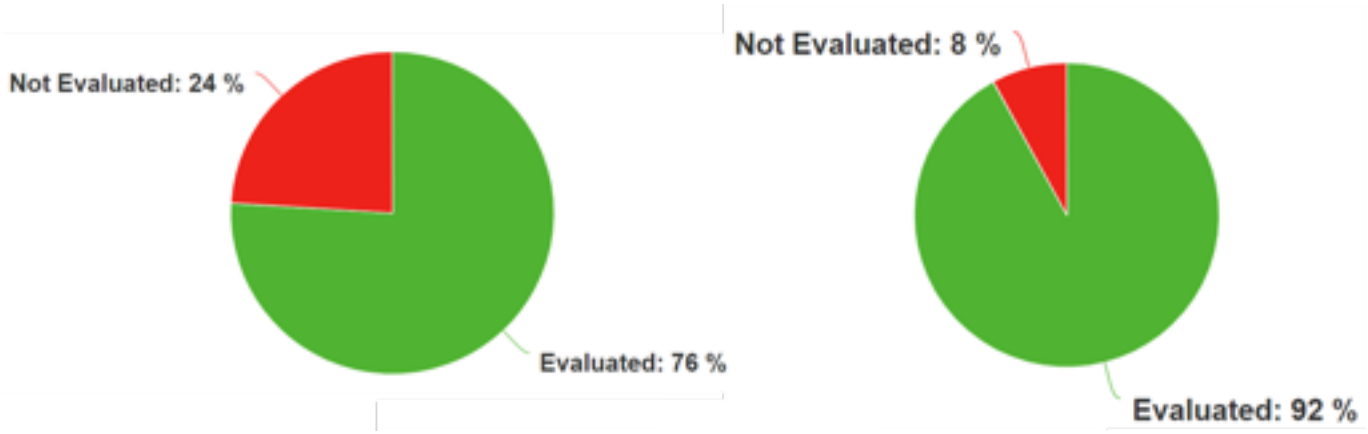


Figure 6.1: On the left side, a pie chart showing the exercises evaluated percentage without considering obstacle occlusions and on the right, considering this phenomenon.

After evaluating the cases of failure, it is necessary to verify the system's performance in real-time using several types of hardware to provide greater insight into how the system performs. For this purpose, three devices will perform the exercise evaluation task with the following specifications described in the table below.

ID	Processor	Processor's Frequency	Cores	RAM
1	Intel® Core™ i7-8700	3.2GHz	6	16 GB
2	Intel® Core™ i9-9900k	3.6GHz	8	64 GB
3	Intel® Core™ i7-8565U	1.8GHz	4	8 GB

Table 6.1: Hardware Specifications of the devices where the real-time performance of the exercise evaluation procedure was evaluated.

With these devices and taking into account the main events occurring during the user's exercise performance, an evaluation of the system can occur, where the analysis considers that the camera provides images at approximately 30 FPS(frames per second). This analysis uses each event to provide insights into the system's resource usage, displayed in this case based on the duration of the process.

ID	Pygame Events (ms)	Images Transformations (ms)	Skeleton (ms)	Grid Formation (ms)	Application's Widgets (ms)	FPS
1	0.02	32.00	0.212	5.8	1.55	23
2	0.01	28.96	0.212	0.61	1.44	29
3	0.37	74.08	0.34	3.79	3.29	12

Table 6.2: Table with the main events occurring when the user performed an exercise and complementary metrics.

6.2 Pilot Study

The pilot study had the primary purpose of assessing the system's usability, more specifically, the usability of the questionnaires and the application with the virtual personal trainer and the exercise evaluation procedure. This study has the primary goal of gathering information to improve the system's efficiency and quality for future more extensive studies that can occur with the population of interest. As the questionnaires were not the main focus of the dissertation, this pilot study was divided into two stages: the usability questionnaire, which had 6 participants, and the application's usability, which had 17 participants. The participants in the questionnaire study were 33% female and 67% male, ranging from 23 to 35 years. The application pilot study had 18% females and 82% males. Both of these distributions are in the figure below.

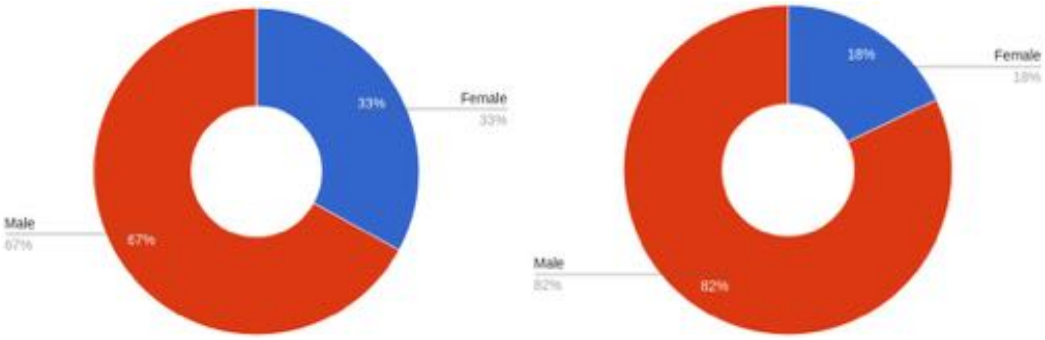


Figure 6.2: Gender Distribution of the Questionnaires(left) and Application(right) usability pilot studies.

The procedure used on these pilot studies is described below:

1. Explain to the participant the main purpose of the experience and how they can interact with the product at hand.
2. Start the questionnaires/application and answer any question the user could have when necessary.

3. After the starting menu, the user can perform the tasks at hand while following the coach's voice lines.

4. When the participant is satisfied with the result of the questionnaires or the number of exercises performed, the participant can perform the User Experience Questionnaire(UEQ).

6.3 Analysis of the Results

The analysis of the results employed a standard User Experience Questionnaire (UEQ). UEQ is a fast and reliable questionnaire to measure the User Experience of interactive products, is available in more than 30 languages and is easy to use due to rich supplementary material[52]. The questionnaire scales cover several usability aspects(e.g., efficiency, perspicuity) and user experience (e.g., originality, stimulation). For a more detailed analysis of these scales, the following figure explains each of these aspects.

Figure 6.3: Explanation of each scale, measured by the User Experience Questionnaire (UEQ)[52].

Within the population from the pilot study, the participants needed to evaluate two primary components of the system in terms of usability, the questionnaires to detect and categorize the user's sarcopenia and the application which allows the user to select and perform exercises appropriate to their condition with the procedures developed on this dissertation. Regarding the questionnaires, the pilot study involved six people, as the questionnaires were not the main focus of this dissertation. Jakob Nielsen et al. Thomas K. Landauer describe a mathematical model to find usability problems[35] showing that the identification of usability problems in a medium-large software project can occur with five test users. Where afterwards, the benefit/cost ratio decreases significantly with the increase of test users. Once the participants finish their experience with the UEQ questionnaire, it

is possible to obtain figure 7.4, which provides the answer distribution of the accumulative results for each item (Y-axis) with their respective percentages of responses (X-axis). From the figure, it is possible to observe a positive reception from the participants in the most relevant items(e.g., ”unfriendly/friendly”, ”confusing/clear”, ”bad/good”, and ”difficult to learn/easy to learn”). From the participant’s answers, it is also possible to extract the mean/standard deviation of the usability aspects (figure 7.5), providing greater insight into the overall system’s evaluation.

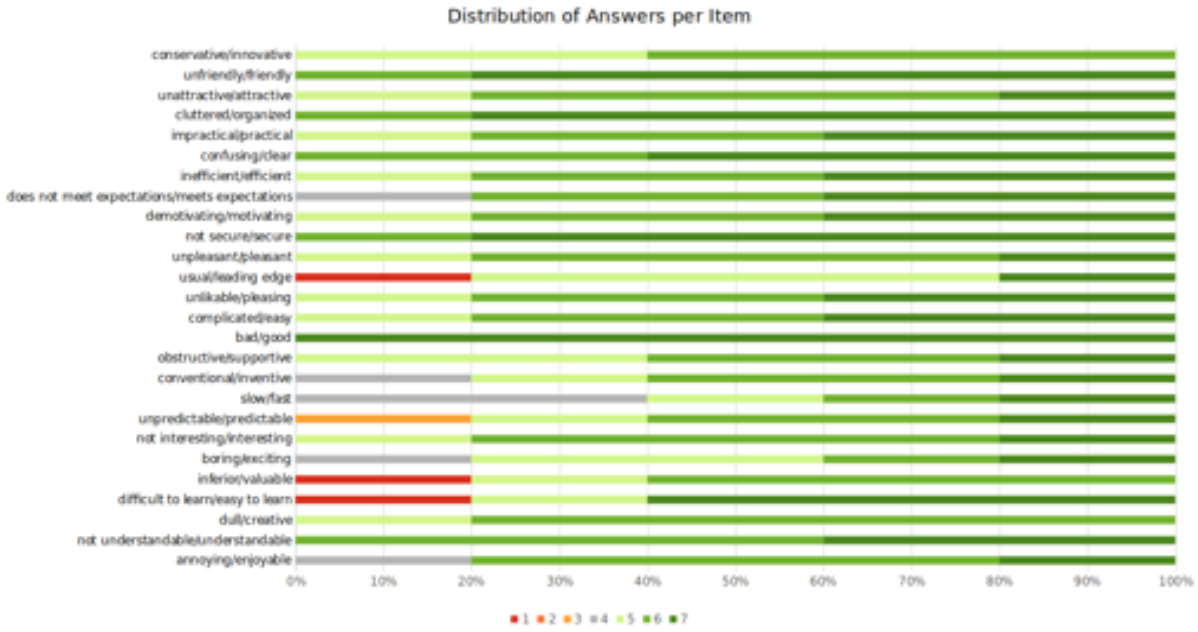


Figure 6.4: Answer Distribution for the Questionnaires experience.

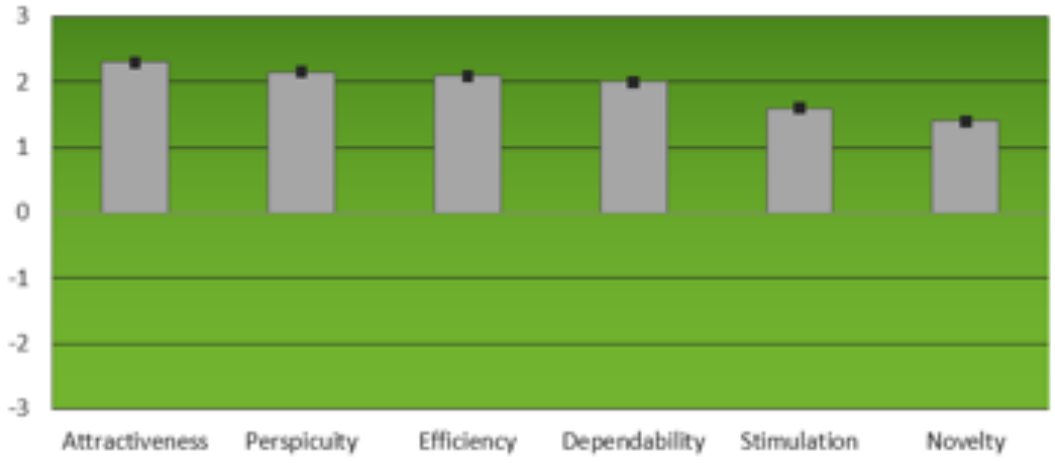


Figure 6.5: Mean Values with Standard Deviation of the Questionnaire’s Usability Scales.

With the systems’ user experience, it is possible to compare the performance with existing values from a reference or benchmark data set that encompasses the evaluation of

different products. The benchmark data set contains data from 21175 persons from 468 studies concerning different products (e.g., business software, web pages, web shops and social networks). The result's comparison of the evaluated product with the data in the benchmark allows conclusions about the relative quality of the product compared to other products. From the figure, it is possible to see that the system's overall performance when compared to others is good, being a valuable tool for the users.

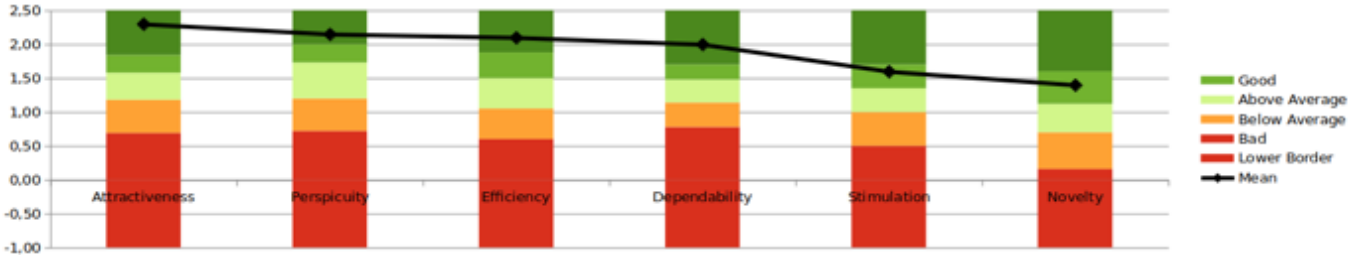


Figure 6.6: Benchmark for the Questionnaires experience.

After the usability evaluation of the questionnaires, the application can be assessed using a subset of the exercises developed in this dissertation. This evaluation provides insights related to the application usability by analyzing with the UEQ the virtual personal trainer interaction, monitorization of the user's performance and the adaptation methods incorporated if the user could not finish the exercise. Similar to the results of the questionnaires, the following figures show good performance in terms of answer distribution, mean/standard deviation of the scales and benchmark. This analysis was possible with the pilot study that had 17 people with a wide range of demographic participating.

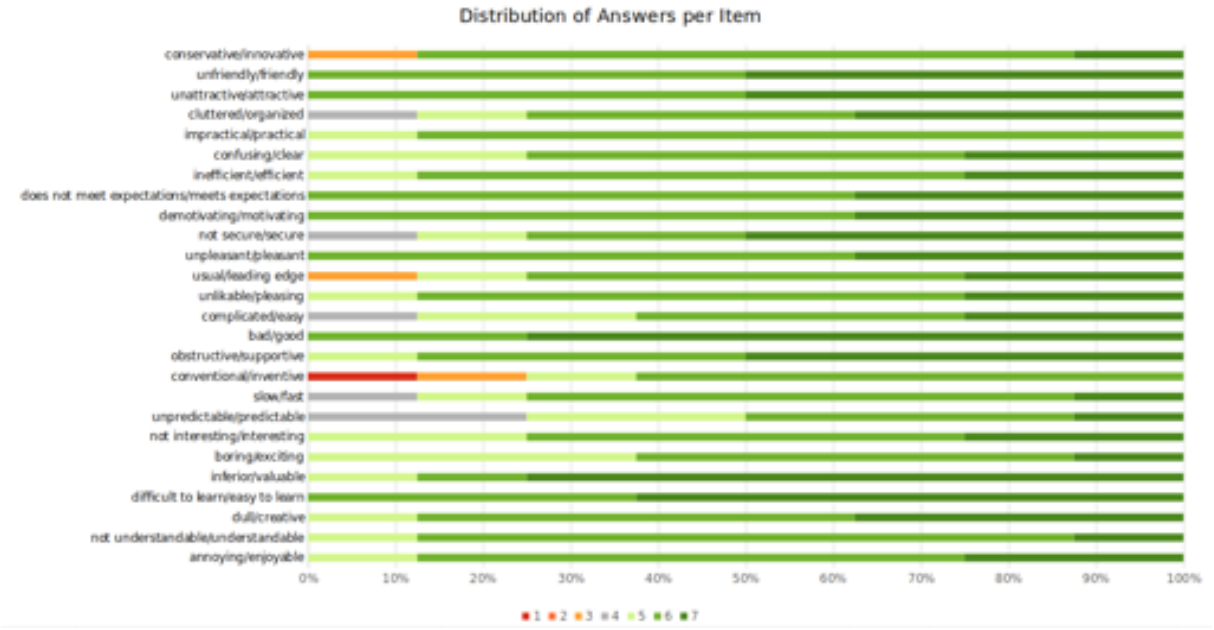


Figure 6.7: Answer Distribution for the Application's experience.

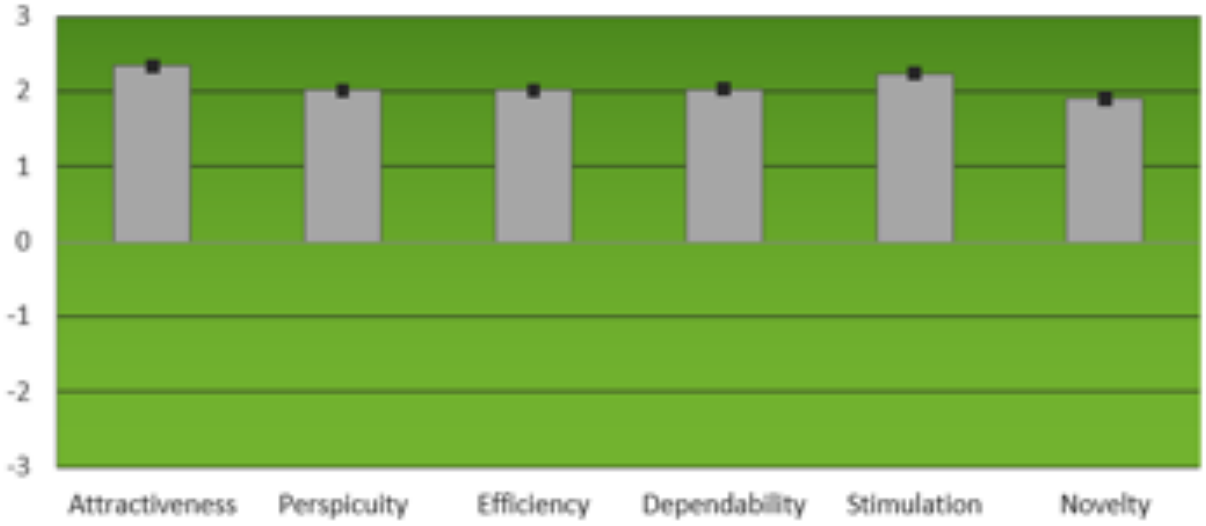


Figure 6.8: Mean Values with Standard Deviation of the Application's Usability Scales.

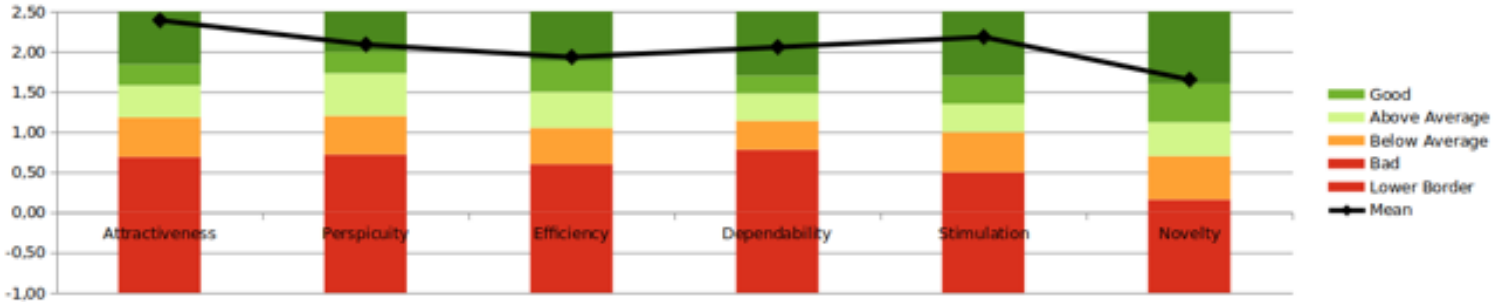


Figure 6.9: Benchmark for the Application experience.

UEQ provides several insights related to usability; however, many aspects regarding individual physical performance need to be assessed to provide a more detailed analysis of the system. The objective of this analysis is to observe the duration of an exercise session depending on the participant's age and the performed exercises. With this objective, there are four major individual parameters:

- Age of the Participant
- Amount of Time
- Amount of Exercises Performed
- Type of the Exercises Performed

The following table displays the user's results for the individual parameters during their exercise sessions while assessing the application's usability.

# User	Age	Time	Amount of Exercises	Exercises
1	26	12 min	4	Seated Opening Arms, Arms Extension w/ Ball, Lifting From Chair and Right Leg Flexion
2	23	11 min	3	Raise Heels And Toes, Seated Opening Arms, Seated Bicep Curl
3	30	11 min	4	Raise Heels And Toes, Seated Opening Arms, Arms Extension w/ Ball and Lifting From Chair
4	31	10 min	3	Seated Opening Arms, Seated Bicep Curl and Seated Chest Press
5	28	7 min	3	Seated Opening Arms, Lifting From Chair and Arms Extension w/ Ball
6	35	8 min	3	Raise Heels And Toes, Seated Opening Arms and Arms Extension w/ Ball
7	24	12 min	3	Seated Opening Arms, Arms Extension w/Ball and Raise Heels And Toes
8	31	13 min	3	Seated Opening Arms, Raise Heels And Toes and Lifting From Chair
9	27	10 min	3	Seated Opening Arms, Raise Heels And Toes and Seated Chest Press
10	21	11 min	3	Seated Opening Arms, Lifting From Chair and Arms Extension w/Ball
11	24	9 min	3	Seated Opening Arms, Arms Extension w/Ball and Lifting From Chair
12	24	10 min	3	Seated Opening Arms, Raise Heels And Toes and Lifting From Chair
13	23	7 min	3	Seated Bicep Curl, Seated Chest Press and Lifting From Chair
14	21	8 min	3	Seated Bicep Curl, Seated Opening Arms and Seated Chest Press
15	21	11 min	3	Seated Opening Arms, Arms Extension w/ Ball and Right Leg Flexion
16	26	12 min	3	Lifting From Chair, Seated Opening Arms and Arms Extension w/ Ball
17	24	9 min	3	Seated Opening Arms, Arms Extension and Lifting From Chair

Table 6.3: Table with individual parameters from their exercise sessions.

To conclude the results chapter, it is essential to mention the presence of the coach's effects on the users. The coach voice lines facilitated the user to navigate the application, provided an interactive experience when performing the questionnaires(e.g., using the voice lines to read the questions) and exemplified how to perform the calibration and the exercise phases. Mainly in the exercise phase, the coach had an impact on the user's exercise performance as the exemplification of the exercise allowed the user to perform a task that they did not know beforehand, eliminating one of the significant barriers to physical activity, lack of knowledge on how to perform the exercise.

6.4 Discussion

The results chapter provided insights into the exercise evaluation procedure and the pilot study. When observing the results of the exercise evaluation, it is possible to observe that only a small amount of the exercises was not possible to characterize due to anomalies of the human pose estimator generated by objects occlusions of the user during the exercises. Exercises such as hip extension had inconsistent trajectories of the landmarks of interest due to occlusions of significant body parts, such as the fact that the human pose estimator relies on to provide an accurate estimation. Furthermore, the raise heels and toes exercise is an example of too subtle changes in the landmarks of interest during the execution, which could not be detected using the number of regions that the system currently uses.

In terms of real-life performance, the results show that the application slows down mainly due to image transformations of the webcam's frames. These transformations were necessary to conform to the main python package for the application's development, pygame, to display and manipulate the image for the final product that the user can see. Furthermore, this package runs on CPU, verified when comparing the performance of the three devices. The one with the best CPU provided the best performance in terms of time duration per task and overall FPS when performing the exercise performance task.

The focus of the pilot studies were to evaluate the usability, interaction and immersive of the system using the User Experience Questionnaire. The first study with 6 participants revealed a positive response from the participants. For example, the mean/standard deviation values of the usability scales (Figure 7.5) are extremely positive, where the standard interpretation of the scales means is that values between -0.8 and 0.8 represent a neutral evaluation, while values above 0.8 represent a positive evaluation and lower then -0.8 represent a negative evaluation[47]. The range of the scales is between -3(bad) and 3(extremely good). In this case, the questionnaire obtained a good performance in attractiveness and efficiency, although it received a lower value in stimulation or novelty, expected due to the nature of the task. The benchmark for this study(figure 7.6) has similar re-

sults where the performance is good but there are user experience qualities which can be improved(e.g., stimulation and novelty).

Regarding the second study, although it had a small sample size of 17 users from a different target population, the results seem optimistic and provide valuable insights to improve the current development. Similar to the questionnaire's analysis, the mean values(figure 7.8) show good product usability, highlighting usability scales such as attractiveness, stimulation and novelty. The application's benchmark received a good response from the participants with attractiveness and stimulation with the highest scores, showing the application's potential to promote physical activity and to be tested with the target population.

When performing the study, one constraint that was not measured was the presence of several people detected by the camera. As mentioned before, the approach taken to detect the human pose is a single-person usage. This fact is important to recall because, during the pilot study, only one person was captured by the camera for each experiment. However, from previous experiments, the pose estimator, when dealing with several people in the image, usually remains with the person closest to the camera, that in most cases, is the one performing the exercise. This occurrence can be an issue for other implementations; however, for this dissertation, since the main goal is for domestic usage, it can be assumed that a small flow of people will be present in the household. Therefore, these occurrences of this happening will be rare, benefiting the current implementation.

From the users, this physical activity promotion was evaluated as appropriate for the elderly population. However, the sarcopenia evaluation, the prescribed exercises and the performance of these need to be assessed by medical personnel as it will bring valuable insights into how the current system can improve when assisting and monitoring the user. One key aspect that can be discussed with physicians is the assessment of users' injuries, which can be helpful for the system so it can provide exercises targeted for the person in terms of sarcopenia severity and physical injuries.

7

Conclusion and Future Work

This final chapter of the dissertation presents the conclusions taken from the previous chapter, results and discussion while also providing suggestions for future work as this dissertation is a project in ActiVas Framework, which is an ongoing project.

7.1 Conclusion

This work had the objective of implementing a virtual personal trainer for active ageing using a 2D generalistic descriptor that could translate the execution of several exercises in an in-house environment using a setup that is affordable to be scalable. The recording of the exercises uses optical motion capture technology and a certified actor that allows motion transfer between a real-life person and a virtual entity, providing a realistic outcome. With these motions, the avatar can display the correct performance of the exercises using a reliable, low-cost video format that allows adaptation in the way the coach exemplifies the exercises. The format allows the system to provide the user with the feeling of interaction between him and the coach. Furthermore, it allows a customizable experience with a low cost in terms of real-time performance since the exemplification of the exercises is stored in the system. The transformation to these comes down to simple frame encoding modifications.

The setup uses a single fixed camera that allows the user to perform exercises in an exercise zone(visible to the setup) defined by him and be evaluated in real-time while providing feedback on their performance with voice cues or visual cues using a colour scheme. For this purpose, the camera needs to be at a height level that resembles the recordings of the personal trainer when he was performing the exercises. With this setup, the system uses pose estimation technology that relies heavily on an ML algorithm to provide 3D normalized coordinates of landmarks that constitute the skeleton of a person using a single-person approach. However, this dissertation's purpose was to simplify the exercise analysis process by lowering the dimension of these coordinates from three to two dimensions. With the conjunction of this simplified coordinate system for the landmarks

and with a grid method formed using the most static anatomical landmarks along time, the system can use a low amount of landmarks (landmarks of interest) and the trajectory of these to characterize a proper execution of the exercise. When the system detects poor execution, the coach notifies the person using a visual cue on the screen to reinforce the feeling of supervision when performing the exercise.

However, when implementing this methodology, some difficulties must be accounted for since they can worsen the overall system's performance. Highly specialized equipment and trained personnel must record the coach's exercise performance when dealing with Optical Motion Capture. Although this type of motion capture technology is the most state-of-the-art and robust technology, it still raises concerns regarding cost and time due to its learning curve.

Another time-consuming task in this work is the motion transfer between the actor and the virtual coach, which requires multiple software to fully integrate the animation into the character and the character into the correct environment to provide the user with realism. When dealing with exercises with objects, occlusion due to objects between the user and the camera affects the performance of the pose estimator, which is a common issue even for state-of-the-art technology. Another type of exercise that the system struggles to evaluate is when the execution of the exercise involves rapid changes over time, which does not specify a generalistic trajectory due to the random behaviour of the pose estimator, which has a prediction with a high jitter associated with it. Another concern regarding this system is the need for human motion analysis by a technician, leading to the need to specify landmarks of interest and their trajectory to allow the system to evaluate other users.

Regardless of the issues mentioned above, when users remain in the exercise zone and perform the exercises correctly, the system is effective when evaluating the performance of 19 exercises.

7.2 Future Work

This dissertation focused on a virtual personal trainer that teaches and guides people's performance when executing exercises that address sarcopenia. With the assistance of a 2D grid method based on anatomical points of the user's skeleton, this algorithm evaluates the motion by comparing it to the coach using an exercise buffer. The evaluation relies on a single camera that supplies a frame input to the pose estimator algorithm, which is the backbone of the exercise analysis procedure. Nevertheless, it also provides adaptation procedures that provide a customizable experience, transforming an ordinarily tedious and demanding task into an interactive experience that can boost motivation.

In future works, it is crucial to use this assessment technique as a baseline for other as-

sessment methods using different hard-coded approaches or approaches that rely on ML techniques that have revolutionized previous problems in the computer vision field. Furthermore, the interaction between the coach and the person can be improved dramatically by using human emotion recognition software, allowing the system to interact with the person based on the emotions that the person is facing. For example, if the person feels sad or frustrated after failing to perform the exercise, the coach can comfort the user and select another prescribed exercise. While monitoring the user, the system can observe the exercise and the person's emotions, rewarding positive emotions such as happiness. This mechanism is possible due to the exercises at hand since these are naturally slow-pace targeted to older people, allowing, in theory, stable facial features to be detected and tracked over time.

One way to improve the current work development is to add other physical activities evaluation regarding exercises and gait or posture issues to assess the presence of other medical conditions expected for the elderly such as Parkinson's disease. With motion capture technology, it is possible to evaluate the person's gaze and posture to provide conclusions regarding the presence of different conditions in real-time using novelistic procedures involving machine learning algorithms with a good percentage of accuracy in the overall results[14]. Using this premise, one possible direction for this project is the pose estimator to provide the features for machine learning algorithms to evaluate the characteristics in real time without an expensive system/equipment. This approach would provide an evaluation method with more freedom regarding the location and the budget necessary for this work.

Another feature that is added to the system is related to nutrition. Nutrition is a highly relevant topic when a person is affected by sarcopenia and in the elderly population. Although the information described in this dissertation is insufficient to provide a state-of-the-art nutritional algorithm, it does provide a general idea of the key concepts that need to be explored in future development.

As mentioned in the chapter regarding the results, the assistance of medical personnel would benefit the application as they could provide valuable insights that are only present when dealing with this type of problem and the target population. This assistance would be helpful in the development but also during/after the usage of the system where future development could store user's information (e.g., amount of exercises performed by the user, improvement over sessions regarding the development of sarcopenia). After storing and processing this data, physicians could view the user's data to prescribe or update the user's treatments or regimens.

To conclude, facial rigging can significantly benefit the coach since it allows the personal trainer to convey emotions. However, this work is highly demanding due to the amount

of time required to accomplish it and the amount of knowledge that the animator needs of the human face when performing different types of phonemes to provide a sensation of realism when the coach is speaking. Regardless, this work would drastically improve the interaction's realism and, therefore, the overall system.

Bibliography

[1] Ana C. Antunes, Daniela A. Araújo, Manuel T. Veríssimo, and Teresa F Amaral. Sarcopenia and hospitalisation costs in older adults: a cross-sectional study. *Nutrition Dietetics*, 74(1):46–50, June 2016.

[2] Valentin Bazarevsky, Ivan Grishchenko, Karthik Raveendran, Tyler Zhu, Fan Zhang, and Matthias Grundmann. Blazepose: On-device real-time body pose tracking. *arXiv preprint arXiv:2006.10204*, 2020.

[3] Valentin Bazarevsky, Yury Kartynnik, Andrey Vakunov, Karthik Raveendran, and Matthias Grundmann. Blazeface: Sub-millisecond neural face detection on mobile gpus. *arXiv preprint arXiv:1907.05047*, 2019.

[4] Zeeshan Bhatti, Asadullah Shah, Ahmad Waqas, and Nadeem Mahmood. Analysis of design principles and requirements for procedural rigging of bipeds and quadrupeds characters with custom manipulators for animation. *arXiv preprint arXiv:1502.06419*, 2015.

[5] Armin Bruderlin and Lance Williams. Motion signal processing. In *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques - SIGGRAPH '95*. ACM Press, 1995.

[6] Motion Builder. Keyframing, 2019.

[7] Motion Builder. Kinematics, 2022.

[8] Joao Carreira, Pulkit Agrawal, Katerina Fragkiadaki, and Jitendra Malik. Human pose estimation with iterative error feedback, 2015.

[9] T. Cederholm, I. Bosaeus, R. Barazzoni, J. Bauer, A. Van Gossum, S. Klek, M. Muscaritoli, I. Nyulasi, J. Ockenga, S.M. Schneider, M.A.E. de van der Schueren, and P. Singer. Diagnostic criteria for malnutrition – an ESPEN consensus statement. *Clinical Nutrition*, 34(3):335–340, June 2015.

[10] Yao-Jen Chang, Wen-Ying Han, and Yu-Chi Tsai. A kinect-based upper limb rehabilitation system to assist people with cerebral palsy. *Research in Developmental Disabilities*, 34(11):3654–3659, 2013.

[11] Ching-Hang Chen and Deva Ramanan. 3d human pose estimation = 2d pose estimation + matching, 2016.

[12] Alfonso J Cruz-Jentoft, Jean Pierre Baeyens, Jürgen M Bauer, Yves Boirie, Tommy Cederholm, Francesco Landi, Finbarr C Martin, Jean-Pierre Michel, Yves Rolland, Stéphane M Schneider, et al. Sarcopenia: European consensus on definition and diagnosisreport of the european working group on sarcopenia in older peoplea. j. cruz-gentoft et al. *Age and ageing*, 39(4):412–423, 2010.

[13] Alfonso J Cruz-Jentoft, Gülistan Bahat, Jürgen Bauer, Yves Boirie, Olivier Bruyère, Tommy Cederholm, Cyrus Cooper, Francesco Landi, Yves Rolland, Avan Aihie Sayer, Stéphane M Schneider, Cornel C Sieber, Eva Topinkova, Maurits Vandewoude, Marjolein Visser, Mauro Zamboni, Ivan Bautmans, Jean-Pierre Baeyens, Matteo Cesari, Antonio Cherubini, John Kanis, Marcello Maggio, Finbarr Martin, Jean-Pierre Michel, Kaisu Pitkala, Jean-Yves Reginster, René Rizzoli, Dolores Sánchez-Rodríguez, and Jos Schols and. Sarcopenia: revised european consensus on definition and diagnosis. *Age and Ageing*, 48(1):16–31, September 2018.

[14] Marta Isabel ASN Ferreira, Fabio Augusto Barbieri, Vinícius Christianini Moreno, Tiago Penedo, and João Manuel RS Tavares. Machine learning models for parkinson's disease detection and stage classification based on spatial-temporal gait parameters. *Gait & Posture*, 2022.

[15] Weidong Geng and Gino Yu. Reuse of motion capture data in animation: A review. In *Computational Science and Its Applications — ICCSA 2003*, pages 620–629. Springer Berlin Heidelberg, 2003.

[16] M. Gleicher and N. Ferrier. Evaluating video-based motion capture. In *Proceedings of Computer Animation 2002 (CA 2002)*. IEEE Comput. Soc.

[17] Michael Gleicher. Retargetting motion to new characters. In *Proceedings of the 25th annual conference on Computer graphics and interactive techniques*, pages 33–42, 1998.

[18] Xiaonan Guo, Jian Liu, and Yingying Chen. Fitcoach: Virtual fitness coach empowered by wearable mobile devices. In *IEEE INFOCOM 2017 - IEEE Conference on Computer Communications*, pages 1–9, 2017.

[19] Mitsuhiro Hayashibe, Alejandro González, and Maxime Tournier. Personalized balance and fall risk visualization with kinect two. In *2020 42nd Annual International*

Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), pages 4863–4866. IEEE, 2020.

[20] Xin Jin, Yuan Yao, Qiliang Jiang, Xingying Huang, Jianyi Zhang, Xiaokun Zhang, and Kejun Zhang. Virtual personal trainer via the kinect sensor. In *2015 IEEE 16th International Conference on Communication Technology (ICCT)*, pages 460–463, 2015.

[21] Aouaidjia Kamel, Bowen Liu, Ping Li, and Bin Sheng. An investigation of 3d human pose estimation for learning tai chi: A human factor perspective. *International Journal of Human–Computer Interaction*, 35(4-5):427–439, November 2018.

[22] Ladislav Kavan. Part i: direct skinning methods and deformation primitives. In *ACM SIGGRAPH*, volume 2014, pages 1–11, 2014.

[23] Zvi Kons, Slava Shechtman, Alexander Sorin, Carmel Rabinovitz, and Ron Hoory. High quality, lightweight and adaptable tts using lpcnet. In *INTERSPEECH*, 2019.

[24] Pradeep Kumar, Rajkumar Saini, Mahendra Yadava, Partha Pratim Roy, Debi Prosad Dogra, and Raman Balasubramanian. Virtual trainer with real-time feedback using kinect sensor. In *2017 IEEE Region 10 Symposium (TENSYMP)*, pages 1–5, 2017.

[25] K. Kurihara, S. Hoshino, K. Yamane, and Y. Nakamura. Optical motion capture system with pan-tilt camera tracking and real time data processing. In *Proceedings 2002 IEEE International Conference on Robotics and Automation (Cat. No.02CH37292)*. IEEE.

[26] Ariel Kwiatkowski, Eduardo Alvarado, Vicky Kalogeiton, C Karen Liu, Julien Pettré, Michiel van de Panne, and Marie-Paule Cani. A survey on reinforcement learning methods in character animation. In *Computer Graphics Forum*, volume 41, pages 613–639. Wiley Online Library, 2022.

[27] Gene S Lee, Andy Lin, Matt Schiller, Scott Peters, Mark McLaughlin, Frank Hanner, and Walt Disney Animation Studios. Enhanced dual quaternion skinning for production use. In *SIGGRAPH Talks*, pages 9–1, 2013.

[28] Christine Mersiana Lukmanto. Is rotoscope true animation? *ULTIMART Jurnal Komunikasi Visual*, 10(1):12–18, March 2018.

[29] Weian Mao, Yongtao Ge, Chunhua Shen, Zhi Tian, Xinlong Wang, and Zhibin Wang. Tfpose: Direct human pose estimation with transformers. *arXiv preprint arXiv:2103.15320*, 2021.

[30] Julieta Martinez, Rayat Hossain, Javier Romero, and James J. Little. A simple yet effective baseline for 3d human pose estimation, 2017.

[31] A J Mayhew, K Amog, S Phillips, G Parise, P D McNicholas, R J de Souza, L Thabane, and P Raina. The prevalence of sarcopenia in community-dwelling older adults, an exploration of differences between studies and within definitions: a systematic review and meta-analyses. *Age and Ageing*, 48(1):48–56, July 2018.

[32] MediaPipe. Blazepose detector - human pose estimation.

[33] MotionBuilder. Motionbuilder constraints.

[34] Lucas Mourot, Ludovic Hoyet, François Le Clerc, François Schnitzler, and Pierre Hellier. A Survey on Deep Learning for Skeleton-Based Human Animation. *Computer Graphics Forum*, pages 1–32, November 2021.

[35] Jakob Nielsen and Thomas K Landauer. A mathematical model of the finding of usability problems. In *Proceedings of the INTERACT'93 and CHI'93 conference on Human factors in computing systems*, pages 206–213, 1993.

[36] Pedro Alves Nogueira. Motion capture fundamentals a critical and comparative analysis on real-world applications. 2012.

[37] Optitrack. Optical motion capture proceedings, 2020.

[38] Masaki Oshita. Motion volume: Visualization of human motion manifolds. In *The 17th International Conference on Virtual-Reality Continuum and its Applications in Industry*. ACM, November 2019.

[39] New Zealand Otago University. Otago exercise program, 2013.

[40] Yuan Peng. Research on dance teaching based on motion capture system. *Mathematical Problems in Engineering*, 2022:1–8, May 2022.

[41] Polhemus. Polhemus liberty, 2015.

[42] Basilio Pueo and Jose Manuel Jimenez-Olmedo. Application of motion capture technology for sport performance analysis. 2017.

[43] Tiago Henrique Ribeiro and Milton Luiz Horn Vieira. Motion capture technology—benefits and challenges. *Int. J. Innov. Res. Technol. Sci. Int. J. Innov. Res. Technol. Sci*, 48(1):2321–1156, 2016.

[44] Roberta E. Rikli and C. Jessie Jones. The reliability and validity of a 6-minute walk test as a measure of physical endurance in older adults. *Journal of Aging and Physical Activity*, 6(4):363–375, October 1998.

[45] Roberta E. Rikli and C. Jessie Jones. Development and validation of a functional fitness test for community-residing older adults. *Journal of Aging and Physical Activity*, 7(2):129–161, April 1999.

[46] Romeo Šajina and Marina Ivašic Kos. Pose estimation, tracking and comparison.

[47] Dr. Martin Schrepp. User experience questionnaire handbook, 2019.

[48] Samsu Sempena, Nur Ulfa Maulidevi, and Peb Ruswono Aryan. Human action recognition using dynamic time warping. In *Proceedings of the 2011 International Conference on Electrical Engineering and Informatics*. IEEE, July 2011.

[49] Liangchen Song, Gang Yu, Junsong Yuan, and Zicheng Liu. Human pose estimation and its application to action recognition: A survey. *Journal of Visual Communication and Image Representation*, 76:103055, 04 2021.

[50] Marionette Studio. Skeletal animation, 2016.

[51] DD Thompson. Aging and sarcopenia. *Journal of Musculoskeletal and Neuronal Interactions*, 7(4):344, 2007.

[52] Team UEQ. User experience questionnaire(ueq), 2018.

[53] Department of Economic United Nations and Population Division Social Affairs. *World Population Ageing 2019: Highlights*. 2019.

[54] Andrey Vakunov, Chuo-Ling Chang, Fan Zhang, George Sung, Matthias Grundmann, and Valentin Bazarevsky. Mediapipe hands: On-device real-time hand tracking. 2020. https://mixedreality.cs.cornell.edu/workshop.

[55] Daniel Vlasic, Rolf Adelsberger, Giovanni Vannucci, John Barnwell, Markus Gross, Wojciech Matusik, and Jovan Popović. Practical motion capture in everyday surroundings. *ACM Transactions on Graphics*, 26(3):35, July 2007.

[56] Ammar Yasser, Doha Tariq, Radwa Samy, Mennat Allah, and Ayman Atia. Smart coaching: Enhancing weightlifting and preventing injuries. *International Journal of Advanced Computer Science and Applications*, 10(7), 2019.

A

Exercises for Prevention and Rehabilitation of Sarcopenia

Functionality	Exercise
Strength	BreastPlate Bicep Curl Bicep Curl Variation Shoulders Tricep Curl Back Exercise Lifting From Chair Plantar Flexor Dorsiflexor Leg Extension
Low	Seated Chest Press With Elastic Band Seated Arms Extension With Ball Seated Bicep Curl With Elastic Band Seated Opening Arms With Elastic Band Get Up and Sit Down From The Chair Hip Abduction With Elastic Band (Sitting) Raise Heels And Toes (Sitting) Buttock Bridge With Ball Leg Flexion With Chair
Elevated/Moderate	Chest Press With Elastic Band Arms Extension With Ball Rowing With Elastic Band Squat With Chair Support Hip Abduction With Elastic Band Hip Extension

Table A.1: The table provides the list of exercises for prevention(e.g., strength exercises) and rehabilitation(Low and Elevated/Moderate Functionality) of sarcopenia.

B

Questionnaires

B.1 SARC-F Questionnaire

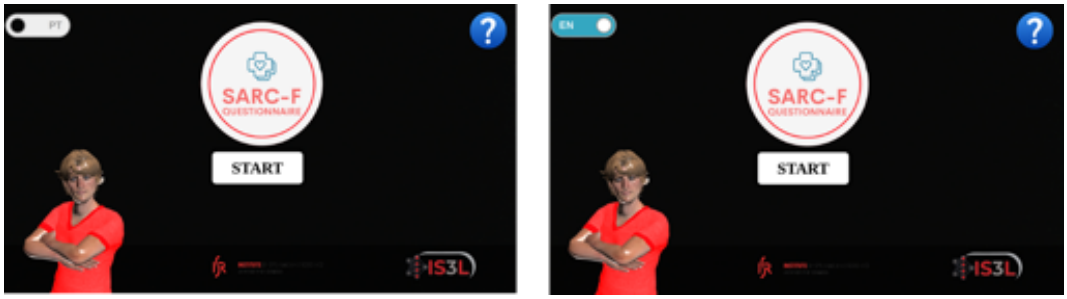


Figure B.1: Inicial Menu from the SARC-F Questionnaire in the available languages.

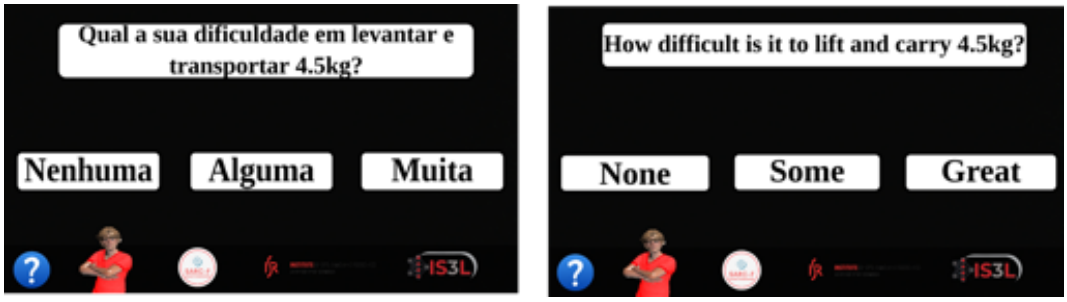


Figure B.2: First Question from the SARC-F Questionnaire in the available languages.

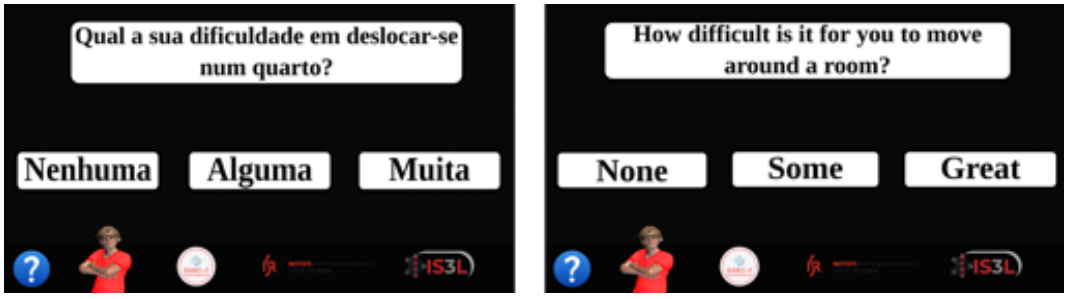


Figure B.3: Second Question from the SARC-F Questionnaire in the available languages.

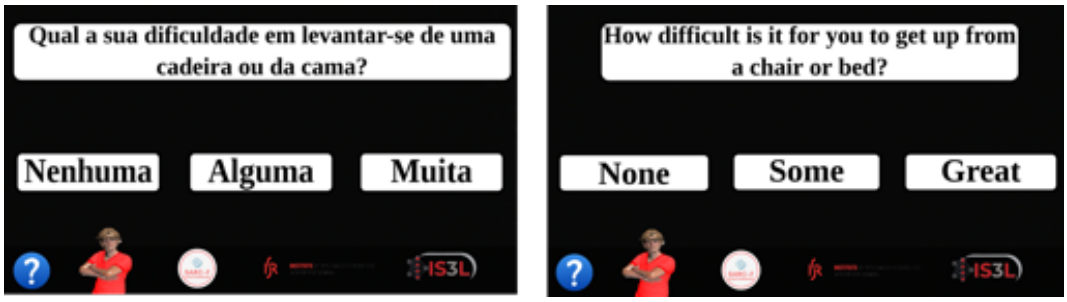


Figure B.4: Third Question from the SARC-F Questionnaire in the available languages.

Figure B.5: Fourth Question from the SARC-F Questionnaire in the available languages.

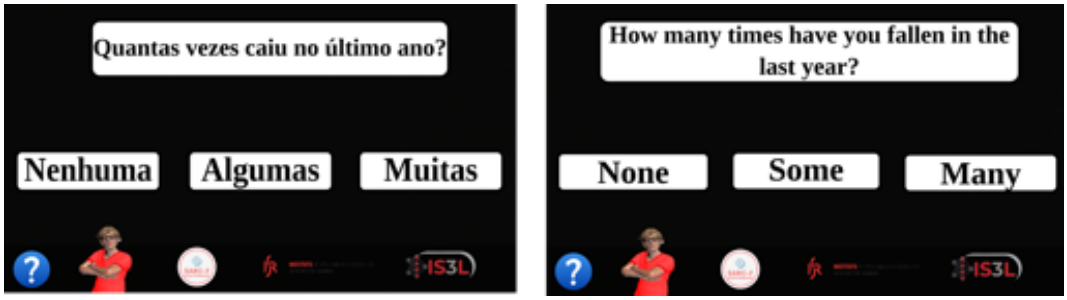


Figure B.6: Fifth Question from the SARC-F Questionnaire in the available languages.

B.2 Rikli-Jones Questionnaire

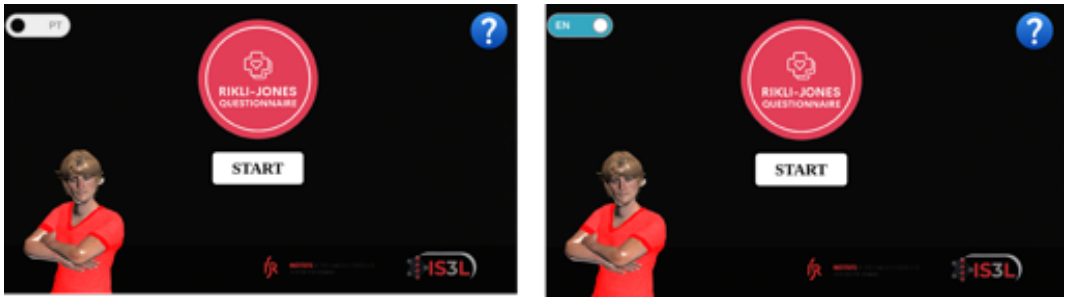


Figure B.7: Inicial Menu from the Rikli-Jones Questionnaire in the available languages.

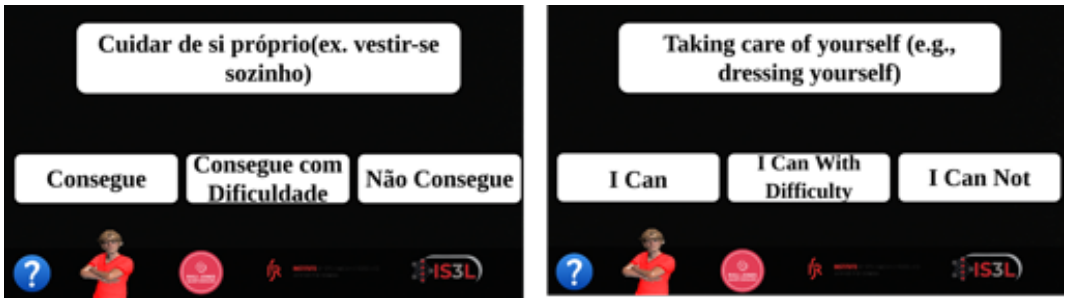


Figure B.8: First Question from the Rikli-Jones Questionnaire in the available languages.

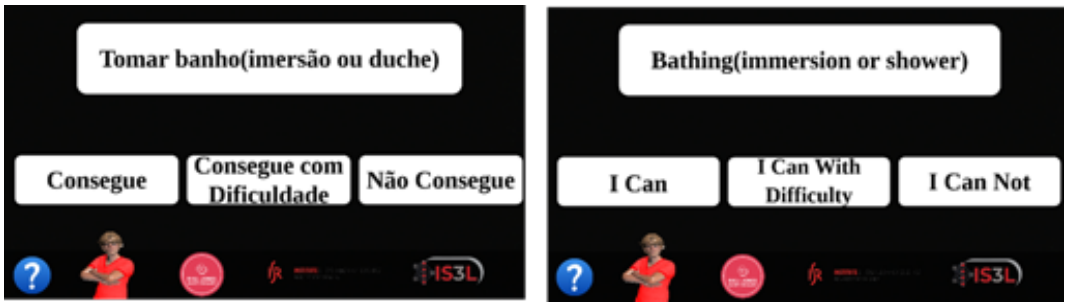


Figure B.9: Second Question from the Rikli-Jones Questionnaire in the available languages.

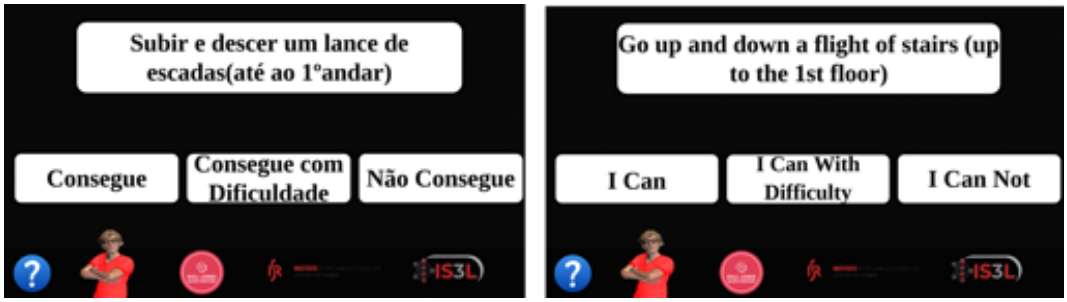


Figure B.10: Third Question from the Rikli-Jones Questionnaire in the available languages.

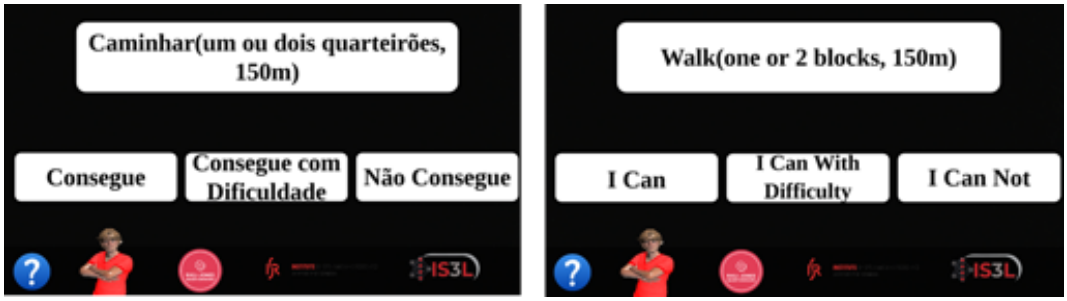


Figure B.11: Fourth Question from the Rikli-Jones Questionnaire in the available languages.

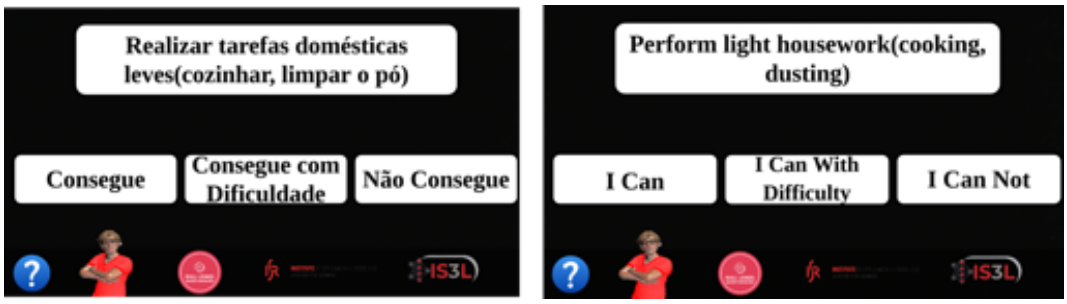


Figure B.12: Fifth Question from the Rikli-Jones Questionnaire in the available languages.

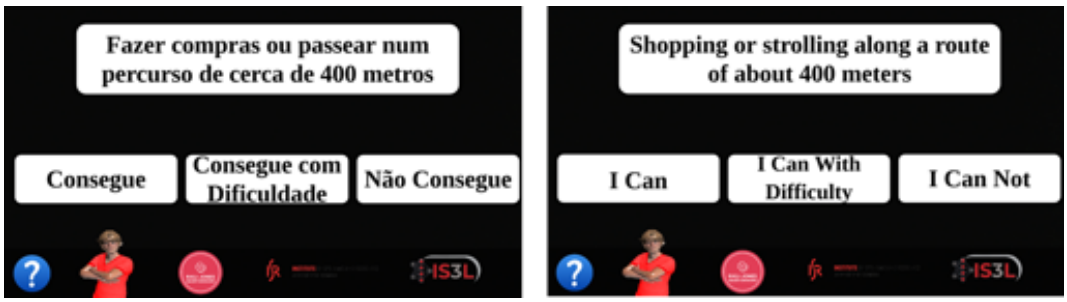


Figure B.13: Sixth Question from the Rikli-Jones Questionnaire in the available languages.

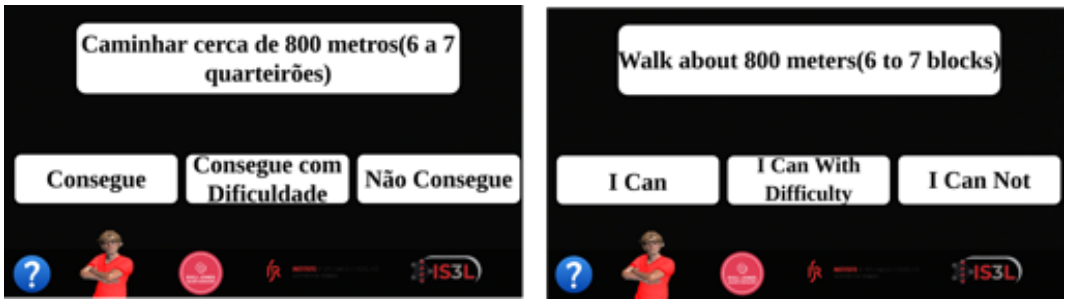


Figure B.14: Seventh Question from the Rikli-Jones Questionnaire in the available languages.

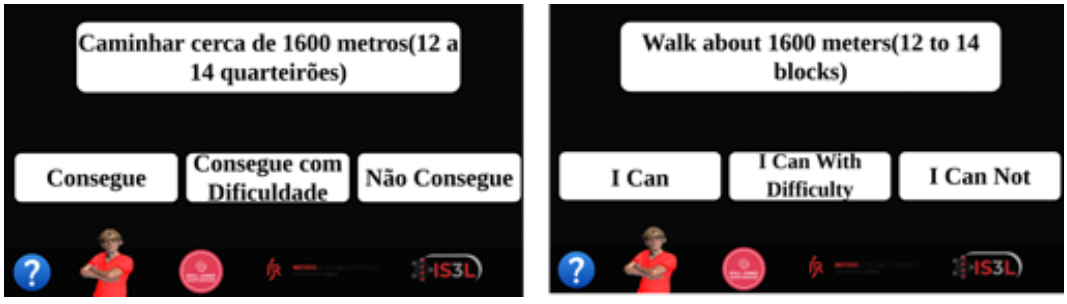


Figure B.15: Eighth Question from the Rikli-Jones Questionnaire in the available languages.

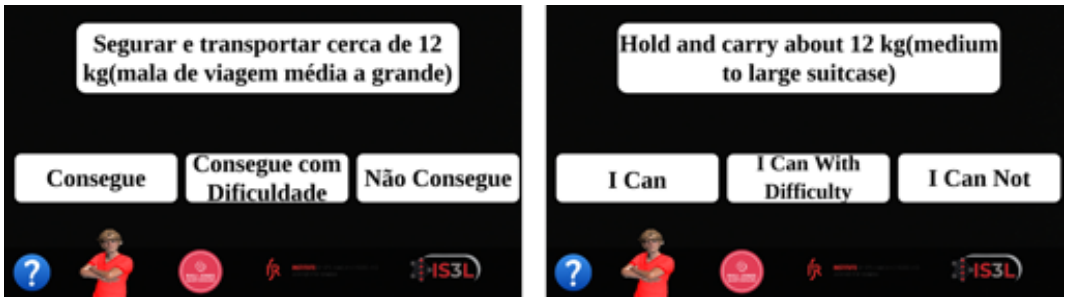


Figure B.16: Ninth Question from the Rikli-Jones Questionnaire in the available languages.

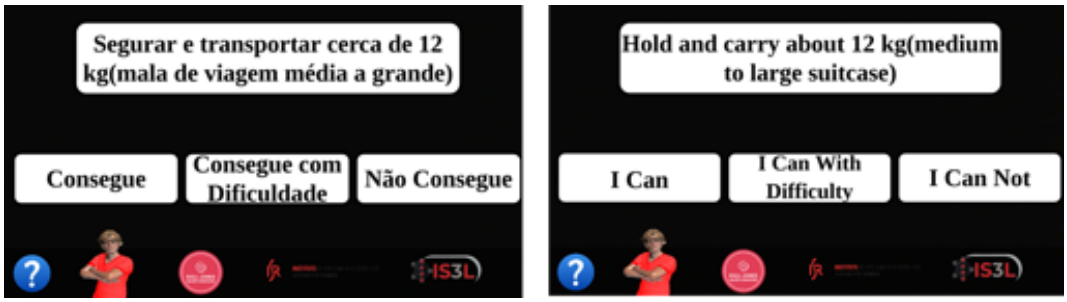


Figure B.17: Tenth Question from the Rikli-Jones Questionnaire in the available languages.

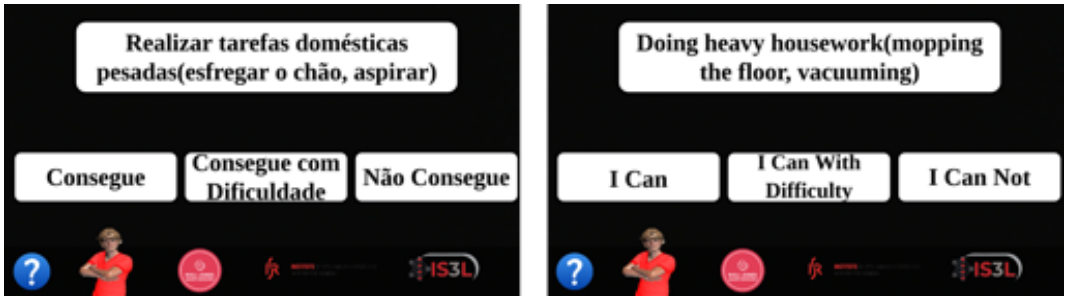


Figure B.18: Eleventh Question from the Rikli-Jones Questionnaire in the available languages.

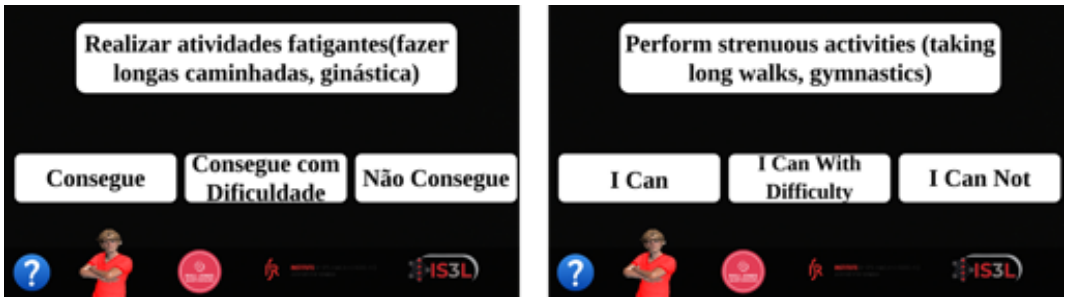


Figure B.19: Twelfth Question from the Rikli-Jones Questionnaire in the available languages.

C

Animation Of Sarcopenia Exercises

C.1 Strength Exercises

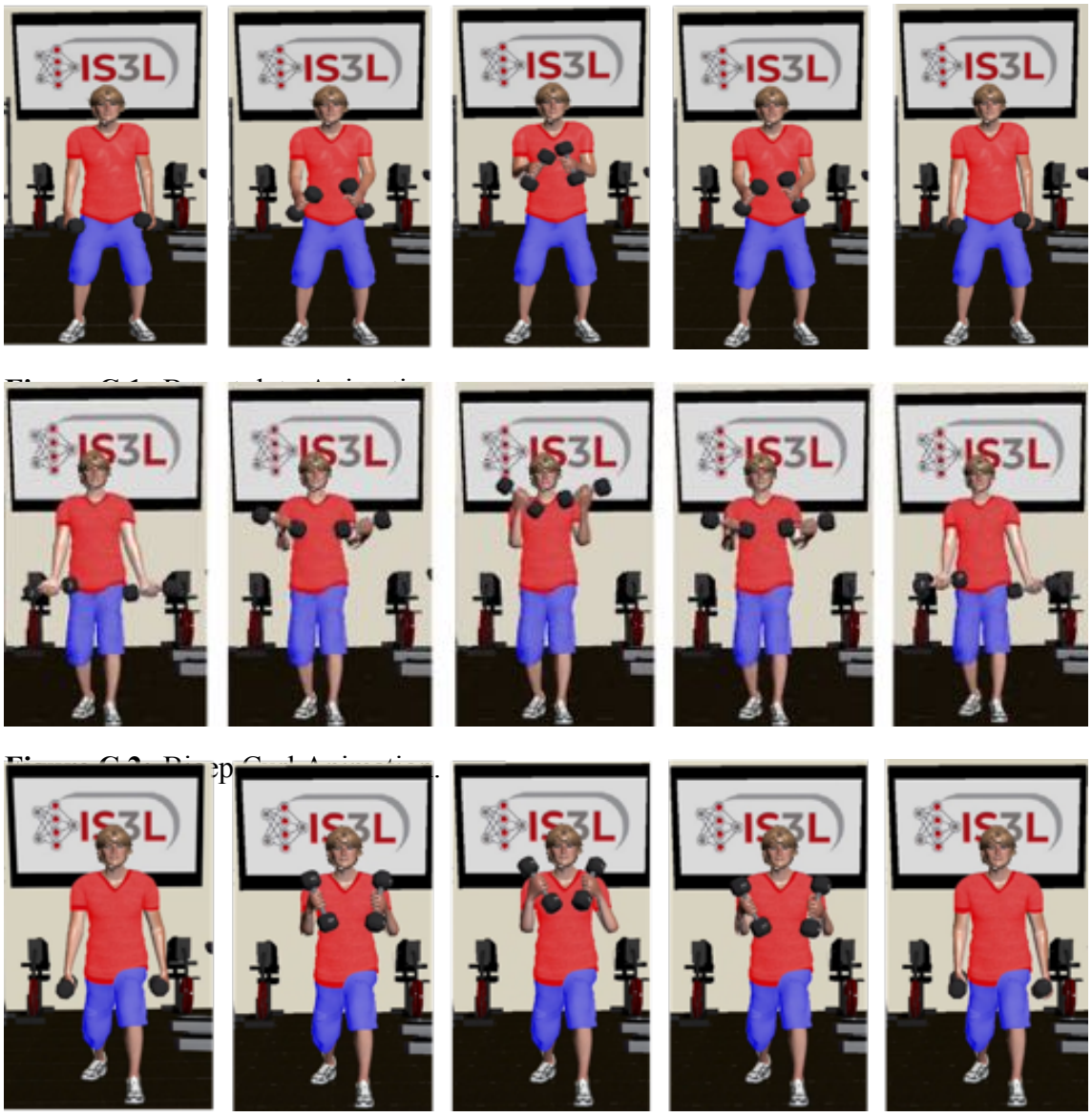


Figure C.3: Bicep Curl Variation Animation.

Figure C.4: Shoulders Exercise Animation.

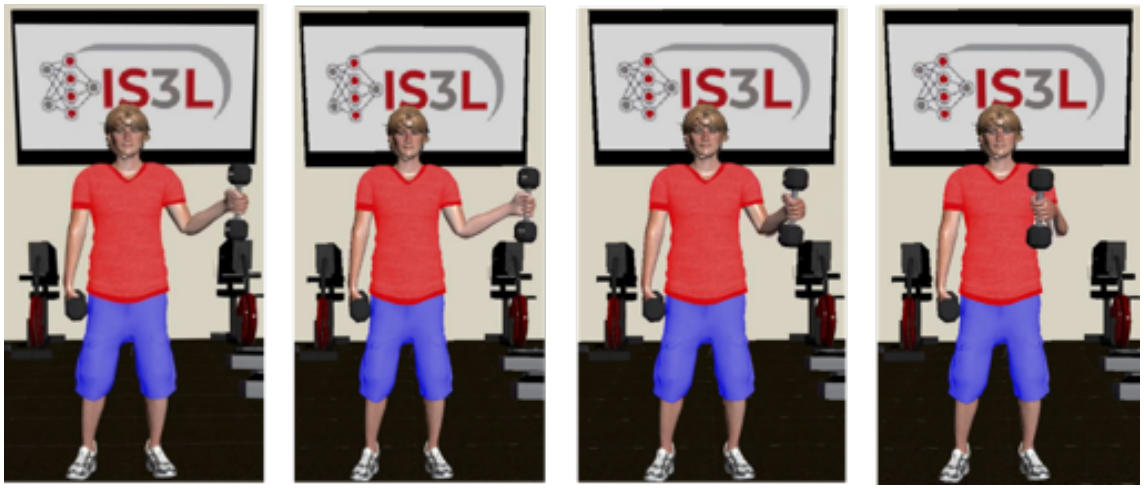


Figure C.5: Left Tricep Exercise Animation.

Figure C.6: Right Tricep Arm Exercise Animation.

Figure C.7: Back Exercise 6 Animation.

Figure C.8: Lifting from Chair Animation.

Figure C.9: Right Plantar Flexor Animation.

Figure C.10: Left Plantar Flexor Animation.

Figure C.11: Left Plantar Dorsiflexor Animation.

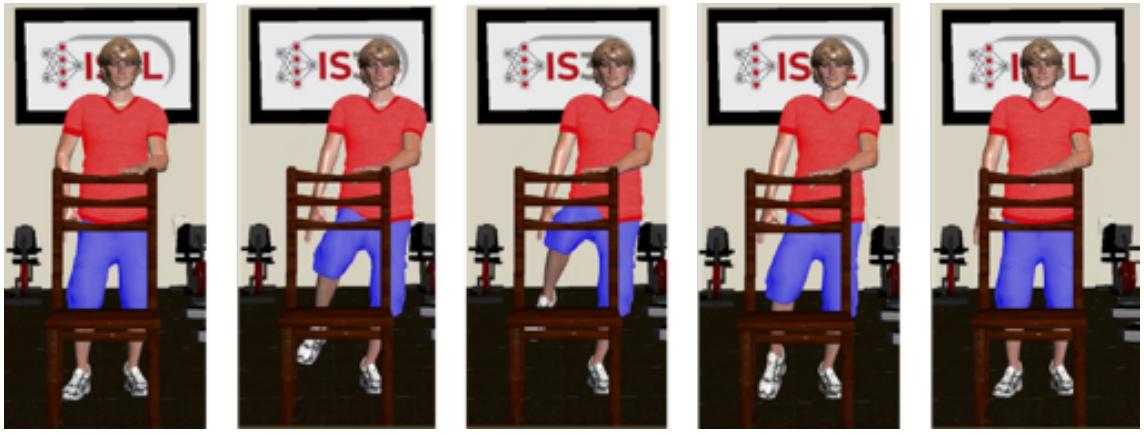


Figure C.12: Right Plantar Dorsiflexor Animation.

Figure C.13: Leg Extension Animation.

C.2 Low Functionality Exercises

Figure C.14: Buttock Bridge with Ball Animation.

Figure C.15: Seated Chest Press Animation.

Figure C.16: Arms Extension with Ball Animation.

Figure C.17: Chair Lift Animation.

Figure C.18: Seated Bicep Curl Animation.

Figure C.19: Hip Abduction Animation.

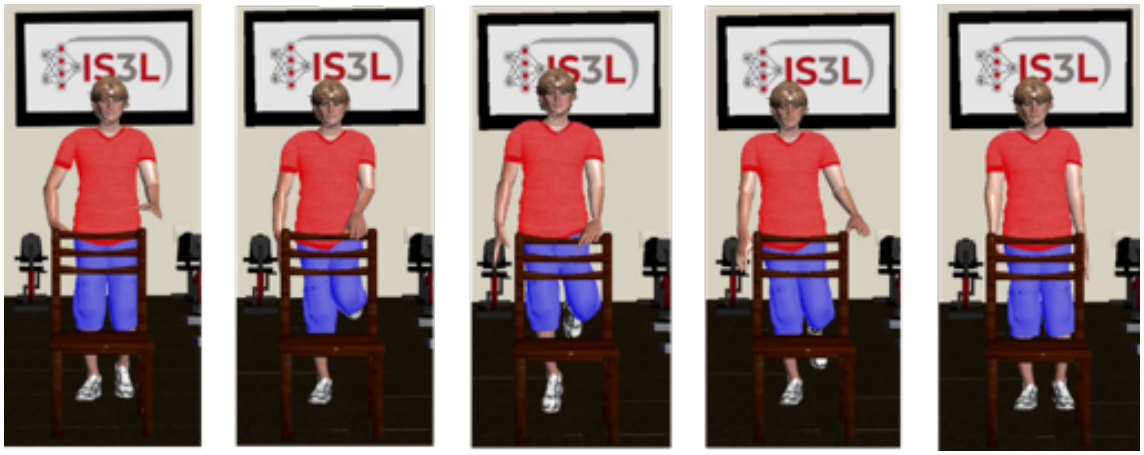


Figure C.20: Leg Flexion Left Leg Animation.

Figure C.21: Leg Flexion Right Leg Animation.

Figure C.22: Raise Heels And Toes Animation.

Figure C.23: Seated Opening Arms Animation.

C.3 Elevated/Moderate Functionality Exercises

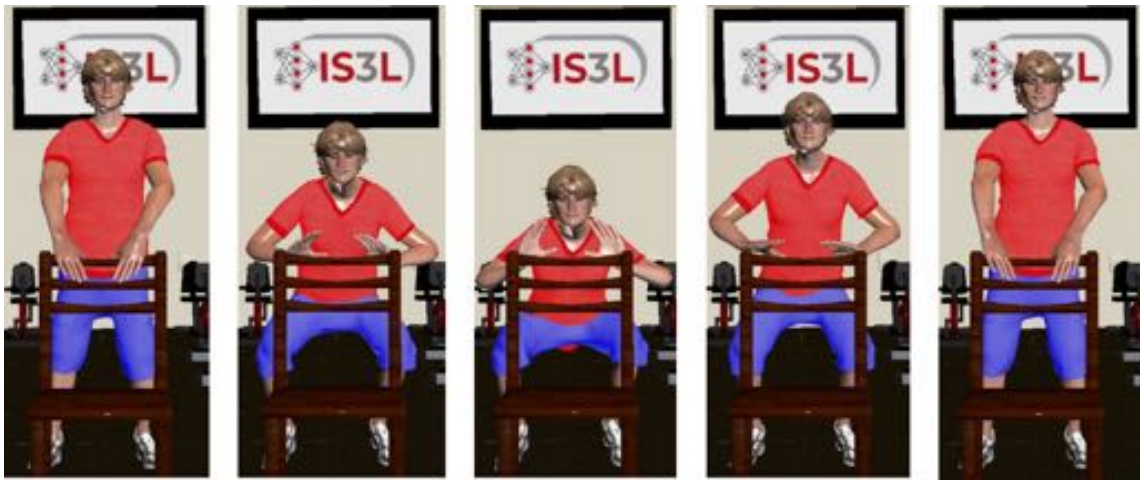


Figure C.24: Squat with Chair Animation.

Figure C.25: Hip Abduction Animation.

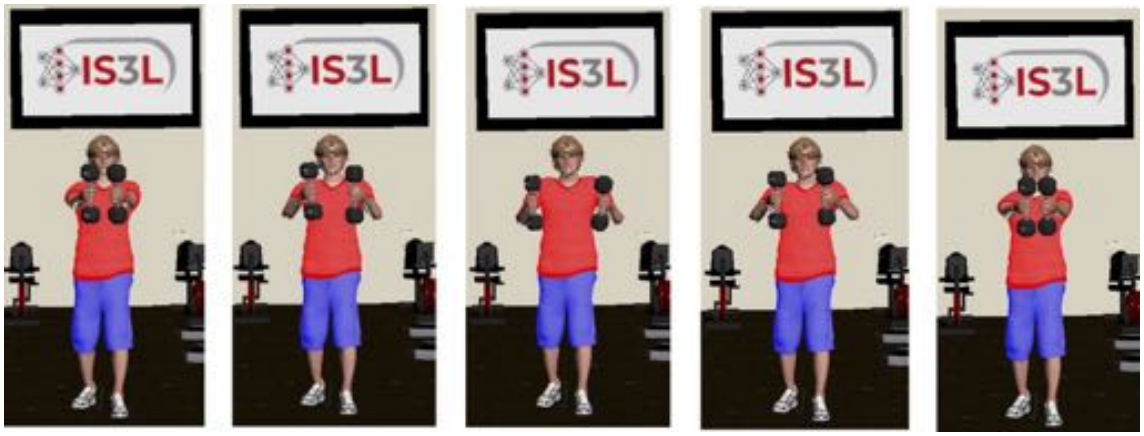


Figure C.26: Paddling Chest Animation.

Figure C.27: Bicep Curl Animation.

Figure C.28: Hip Extension Left Leg Animation.

Figure C.29: Hip Extension Right Leg Animation.

Figure C.30: Standing Chest Press Animation.

Figure C.31: Raise Heels And Toes Animation.

D

SPPB Test

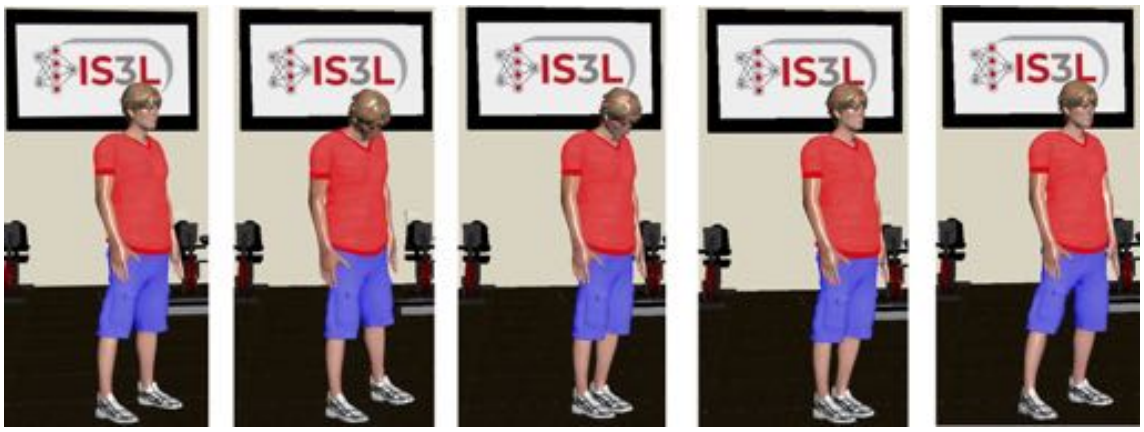


Figure D.1: SPPB Test Animation.

E

TimeStamps Of Sarcopenia Exercises

E.1 Strength Exercises

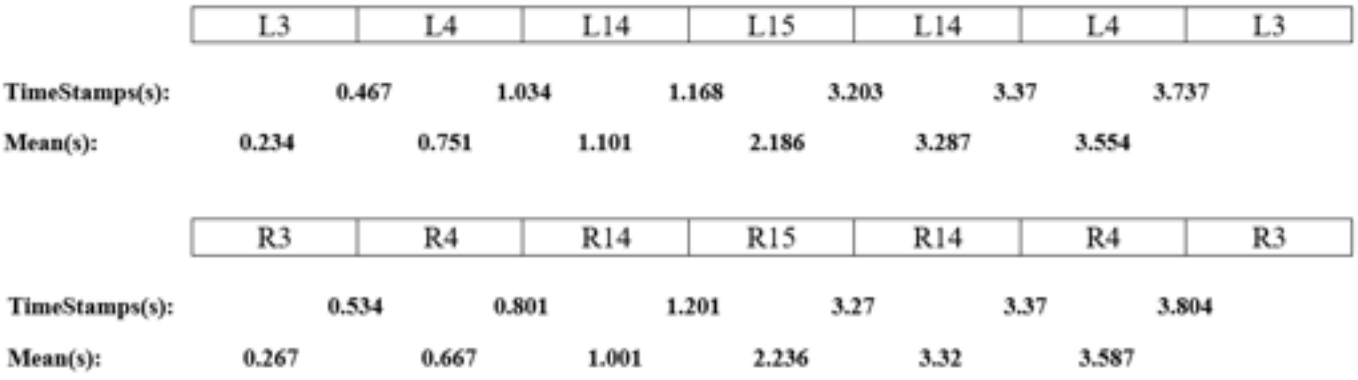


Figure E.1: Breastplate Transitions and Mean TimeStamps.

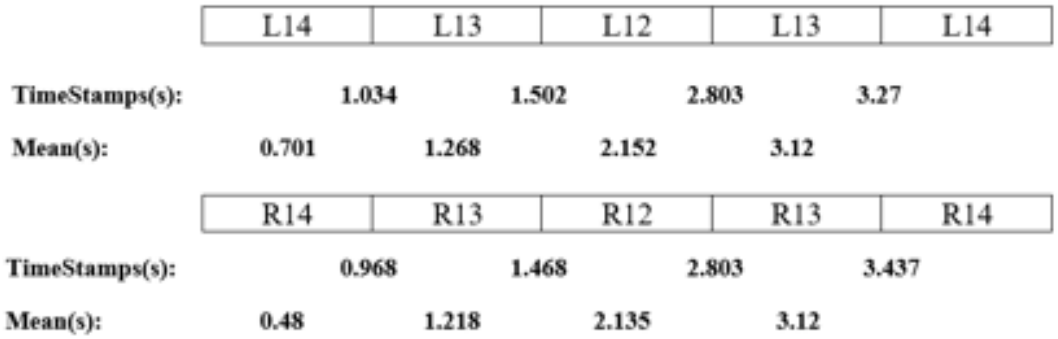


Figure E.2: Bicep Curl Transitions and Mean TimeStamps.

L15	L14	L13	L14	L15
TimeStamps(s):	**0.701**	**1.235**	**3.203**	**3.904**
Mean(s):	**0.350**	**0.968**	**3.203**	**3.554**

R15	R14	R13	R14	R15
TimeStamps(s):	**0.634**	**1.201**	**3.136**	**3.871**
Mean(s):	**0.317**	**0.918**	**2.169**	**3.504**

Figure E.3: Bicep Curl Variation Transitions and Mean TimeStamps.

L14	L24	L14
TimeStamps(s):	**0.801**	**3.77**
Mean(s):	**0.400**	**2.286**

R14	R24	R14
TimeStamps(s):	**0.767**	**3.77**
Mean(s):	**0.384**	**2.402**

Figure E.4: Shoulders Exercise Transitions and Mean TimeStamps.

L13	L23	L33	L23	L13
TimeStamps(s):	**2.236**	**2.903**	**3.67**	**4.304**
Mean(s):	**1.535**	**2.569**	**3.287**	**3.987**

Figure E.5: Left Tricep Curl Exercise Transitions and Mean TimeStamps.

R13	R23	R33	R23	R13
TimeStamps(s):	**0.734**	**1.235**	**2.436**	**3.003**
Mean(s):	**0.367**	**0.984**	**1.835**	**2.719**

Figure E.6: Right Tricep Curl Exercise Transitions and Mean TimeStamps.

L18	L17	L18

TimeStamps(s): **1.331** **2.297**

Mean(s): **0.999** **1.814**

R18	R17	R18

TimeStamps(s): **1.331** **2.363**

Mean(s): **0.849** **1.847**

Figure E.7: Back Exercise Transitions and Mean TimeStamps.

L6	L5	L6

TimeStamps(s): **2.87** **5.772**

Mean(s): **1.585** **4.321**

R6	R5	R6

TimeStamps(s): **2.87** **5.806**

Mean(s): **1.585** **4.338**

Figure E.8: Lifting From Chair Transitions and Mean TimeStamps.

L7	L6	L5	L6	L7

TimeStamps(s): **0.801** **2.336** **2.402** **2.936**

Mean(s): **0.400** **2.302** **1.702** **2.669**

Figure E.9: Left Plantar Dorsiflexor Transitions and Mean TimeStamps.

L7	L6	L5	L6	L7

TimeStamps(s): **0.801** **2.336** **2.402** **2.936**

Mean(s): **0.400** **2.302** **1.702** **2.669**

Figure E.10: Right Plantar Dorsiflexor Transitions and Mean TimeStamps.

L10	L9	L10

TimeStamps(s): **0.767** **4.204**

Mean(s): **0.384** **2.486**

Figure E.11: Left Leg Extension Transitions and Mean TimeStamps.

E.2 Lower Functionality Exercises

L6	L5	L6

TimeStamps(s): **2.936** **5.672**

Mean(s): **1.585** **4.304**

R6	R5	R6

TimeStamps(s): **2.936** **5.672**

Mean(s): **1.585** **4.304**

Figure E.12: Lifting from Chair Transitions and Mean TimeStamps.

Figure E.13: Seated Opening Arms Transitions and Mean TimeStamps.

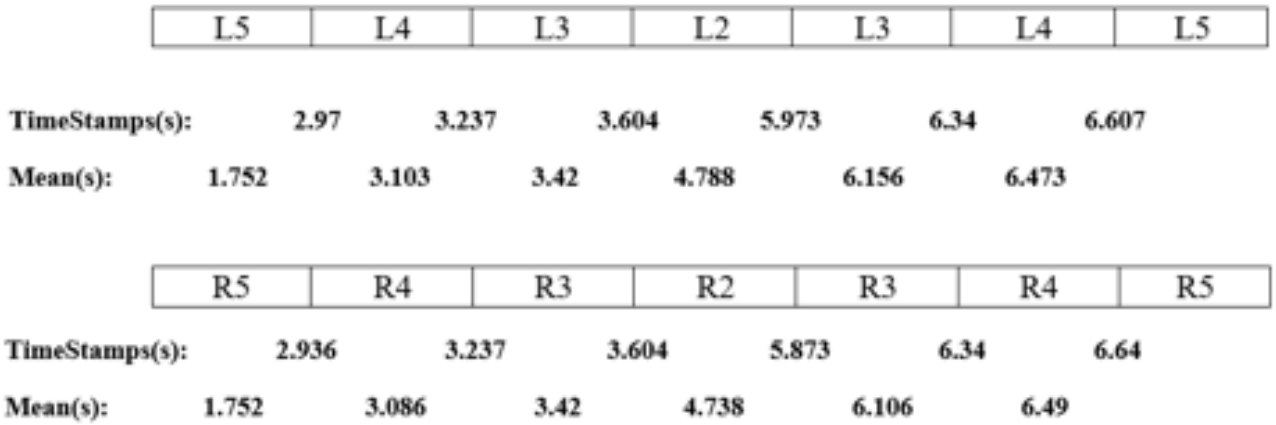


Figure E.14: Seated Arms Extension with Ball Transitions and Mean TimeStamps.

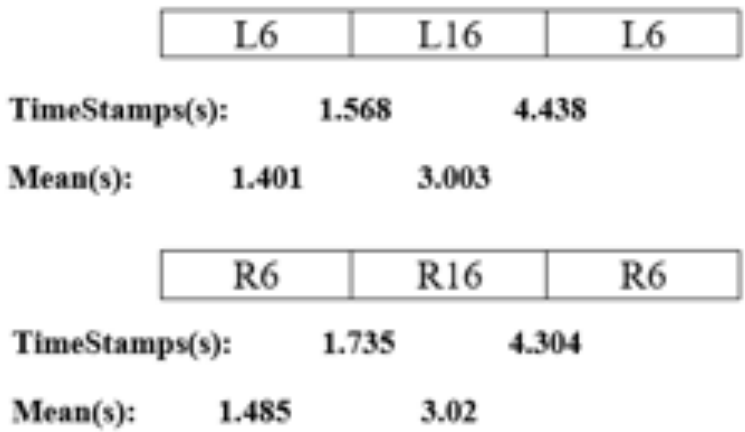


Figure E.15: Hip Abduction Transitions and Mean TimeStamps.

R5	R4	R5

TimeStamps(s): **1.835** **2.936**

Mean(s): **1.602** **2.386**

Figure E.16: Left Leg Flexion Transitions and Mean TimeStamps.

L5	L4	L5

TimeStamps(s): **2.769** **4.605**

Mean(s): **2.569** **3.687**

Figure E.17: Right Leg Flexion Transitions and Mean TimeStamps.

L9	L8	L9

TimeStamps(s): **3.871** **8.709**

Mean(s): **3.253** **6.29**

Figure E.18: Buttock Bridge Transitions and Mean TimeStamps.

L16	L15	L14	L4	L14	L15	L16

TimeStamps(s): **1.134** **1.435** **1.969** **2.336** **3.103** **3.537**

Mean(s): **0.734** **1.285** **1.702** **2.152** **2.719** **3.32**

R16	R15	R14	R4	R14	R15	R16

TimeStamps(s): **1.034** **1.435** **1.902** **2.469** **3.103** **3.67**

Mean(s): **0.684** **1.235** **1.668** **2.186** **2.786** **3.387**

Figure E.19: Seated Bicep Curl Transitions and Mean TimeStamps.

E.3 Elevated/Moderate Functionality

L14	L13	L3	L13	L14

TimeStamps(s): 0.834 1.335 2.436 2.836

Mean(s): 0.417 1.084 1.885 2.636

R14	R13	R3	R13	R14

TimeStamps(s): 0.834 1.335 2.202 2.769

Mean(s): 0.417 1.084 1.768 2.486

Figure E.20: Bicep Curl Transitions and Mean TimeStamps.

L3	L13	L3

TimeStamps(s): 1.268 4.171

Mean(s): 0.868 2.719

R3	R13	R3

TimeStamps(s): 1.168 4.404

Mean(s): 0.851 2.786

Figure E.21: Paddling Chest Transitions and Mean TimeStamps.

L20	L30	L40	L30	L20

TimeStamps(s): 7.574 8.342 9.943 10.477

Mean(s): 3.787 7.958 9.142 10.21

R20	R30	R40	R30	R20

TimeStamps(s): 1.768 2.302 5.072 5.472

Mean(s): 0.884 2.035 3.687 5.272

Figure E.22: Hip Abduction Transitions and Mean TimeStamps.

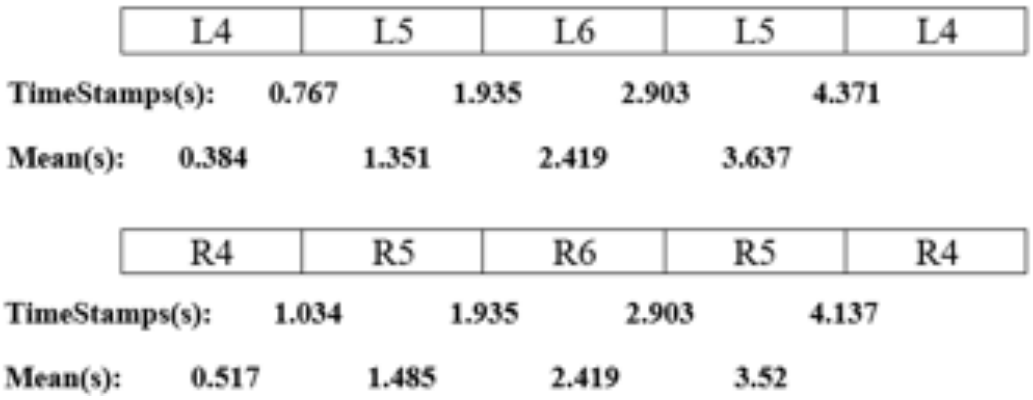


Figure E.23: Squat With Chair Transitions and Mean TimeStamps.

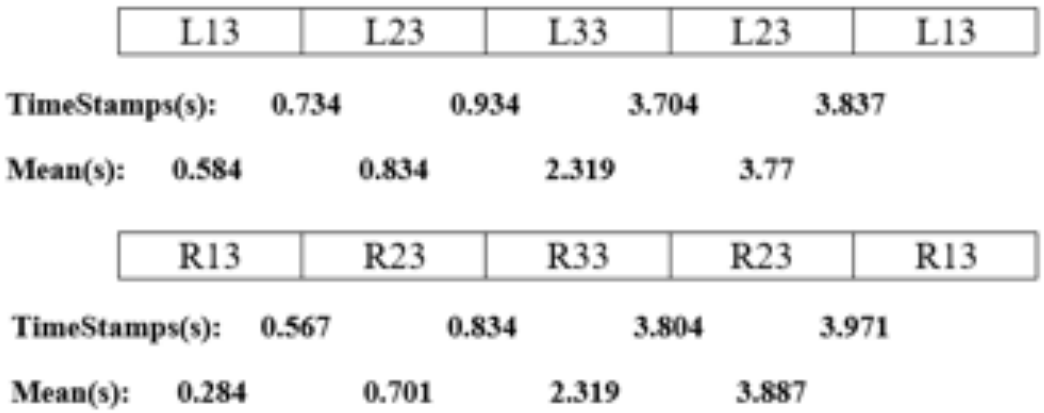


Figure E.24: Standing Chest Press Transitions and Mean TimeStamps.

F

Application Voice Lines

Stage	Voice Line
Inicial Menu	-Hi my name is John and i'm going to be your personal trainer for today, you can select the start option to begin.
Exercise Selection	-**Strength Exercises:** I can see that you have performed your tests and you don't have presence of sarcopenia, let's select one of these exercises and keep it that way. -**Low Functionality:** Based on the results, i can see that the low category is the best one for you, pick one of these exercises to improve your condition. -**Elevated_Moderate Functionality:** Based on the results, i can see that you have elevated or moderate functionality, let's keep it that way and select one of these exercises.
Calibration Phase	-I can see that you have selected the exercise. Now we need to perform the system calibration so it can evaluate you. Put your hands and feet on the red regions and wait 10 seconds.
Show Exercise	-I'm going to show you how you can perform the exercise. Okay? Here we go.
Perform Exercise	-Your turn. You got this, good luck. -Good Job -Keep Going, you are almost there.
Finish Exercise	-You did it, good job!
Didn't Finish Exercise	-Don't worry, we are going to start over with lower weigths and perform the exercise slowly.
After Finishing Exercise	-You can now choose with one hand if you want to perform another exercise or leave.
Adapt Exercise	-**First Time:** Don't worry, this exercise is hard, let's watch the exercise again in a slower version. -**Second Time:** I can see that you haven't been able to complete this exercise. Don't worry, we are going to try again together.

G

Exercise Evaluation - Cases of Failure

G.1 Subtle Changes

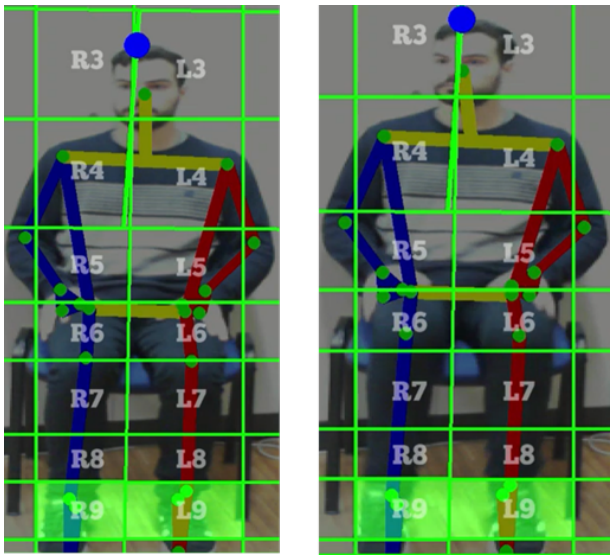


Figure G.1: Raise Heels and Toes from the Low Functionality Exercises.

G.2 Unpredictable Trajectories

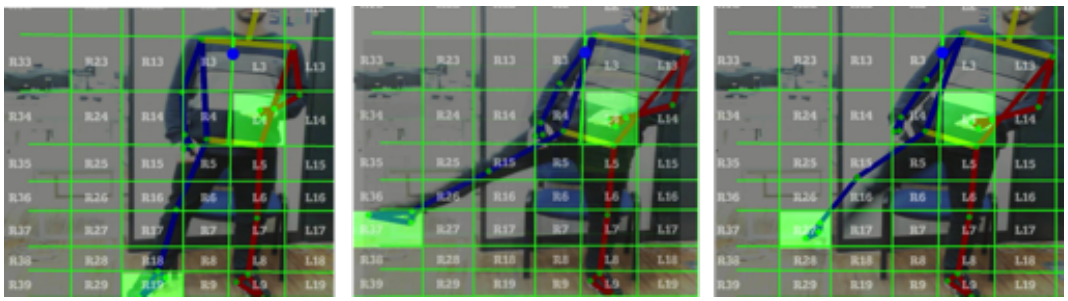


Figure G.2: Correct Right Plantar Flexor repetition from the Strength Exercises.

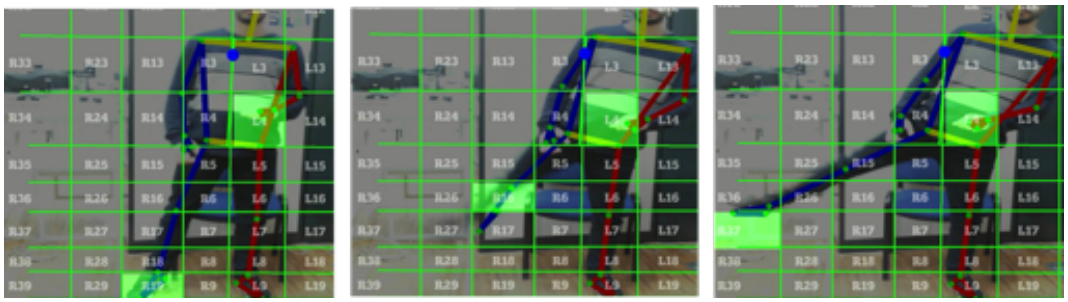


Figure G.3: Wrong Right Plantar Flexor repetition from the Strength Exercises.

Figure G.4: Hip Extension repetition from the Elevated/Moderate Functionality Exercises.

Figure G.5: Different Hip Extension repetition from the Elevated/Moderate Functionality Exercises.

G.3 Obstacles Occlusions

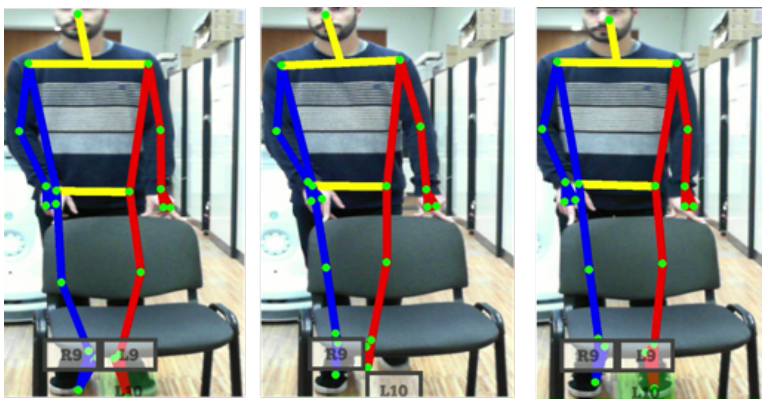


Figure G.6: Leg Extension repetition from the Strength Exercises without obstacles.

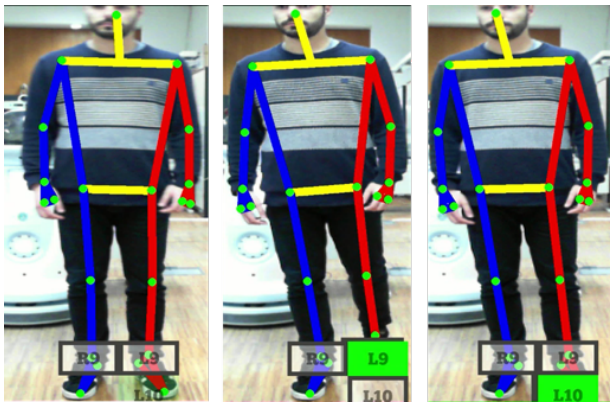


Figure G.7: Leg Extension repetition from the Strength Exercises with obstacles.

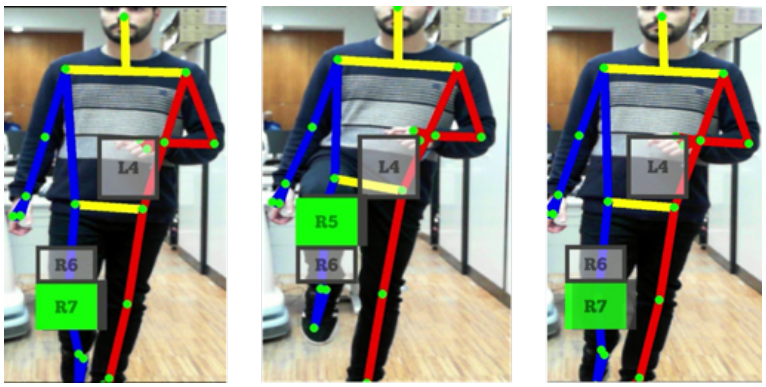


Figure G.8: Leg Flexion repetition from the Low Functionality Exercises without obstacles.

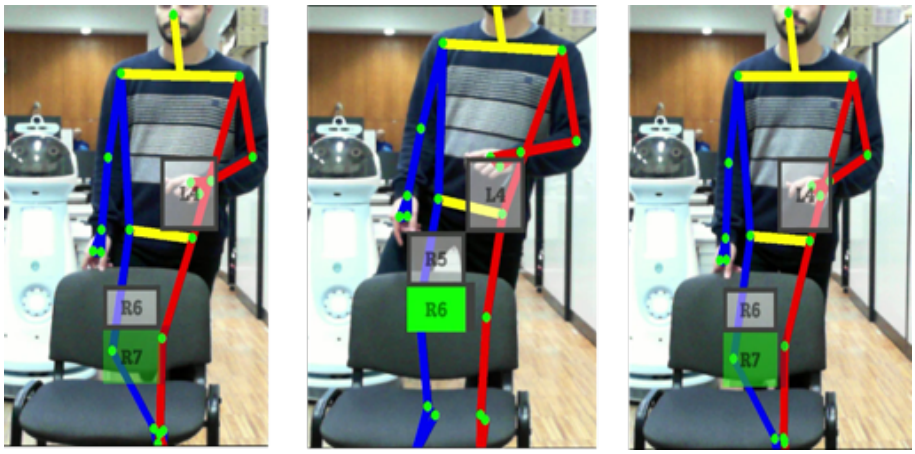


Figure G.9: Leg Flexion repetition from the Low Functionality Exercises with obstacles.

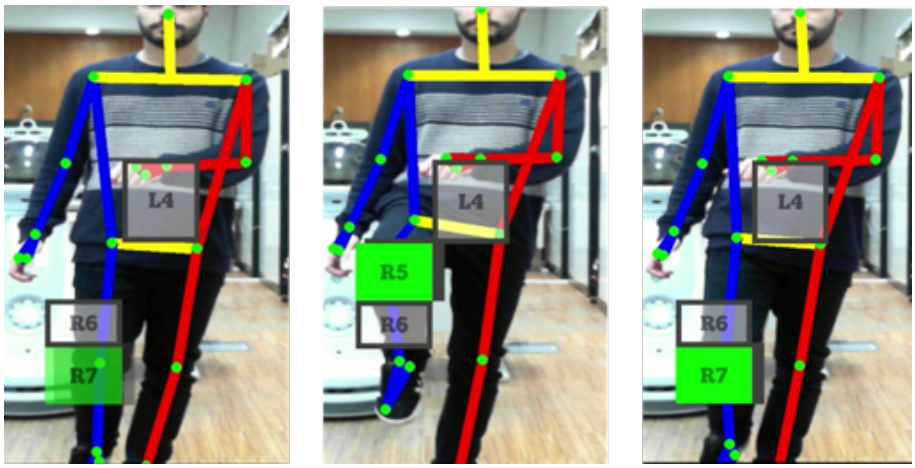


Figure G.10: Right Plantar Flexor repetition from the Strength Exercises without obstacles.

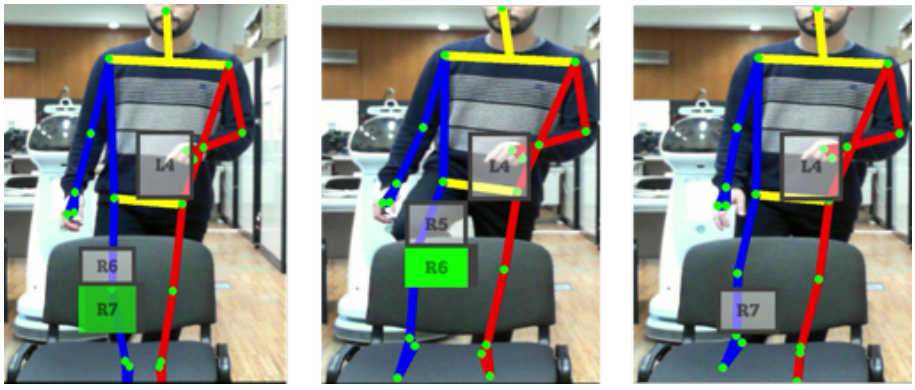
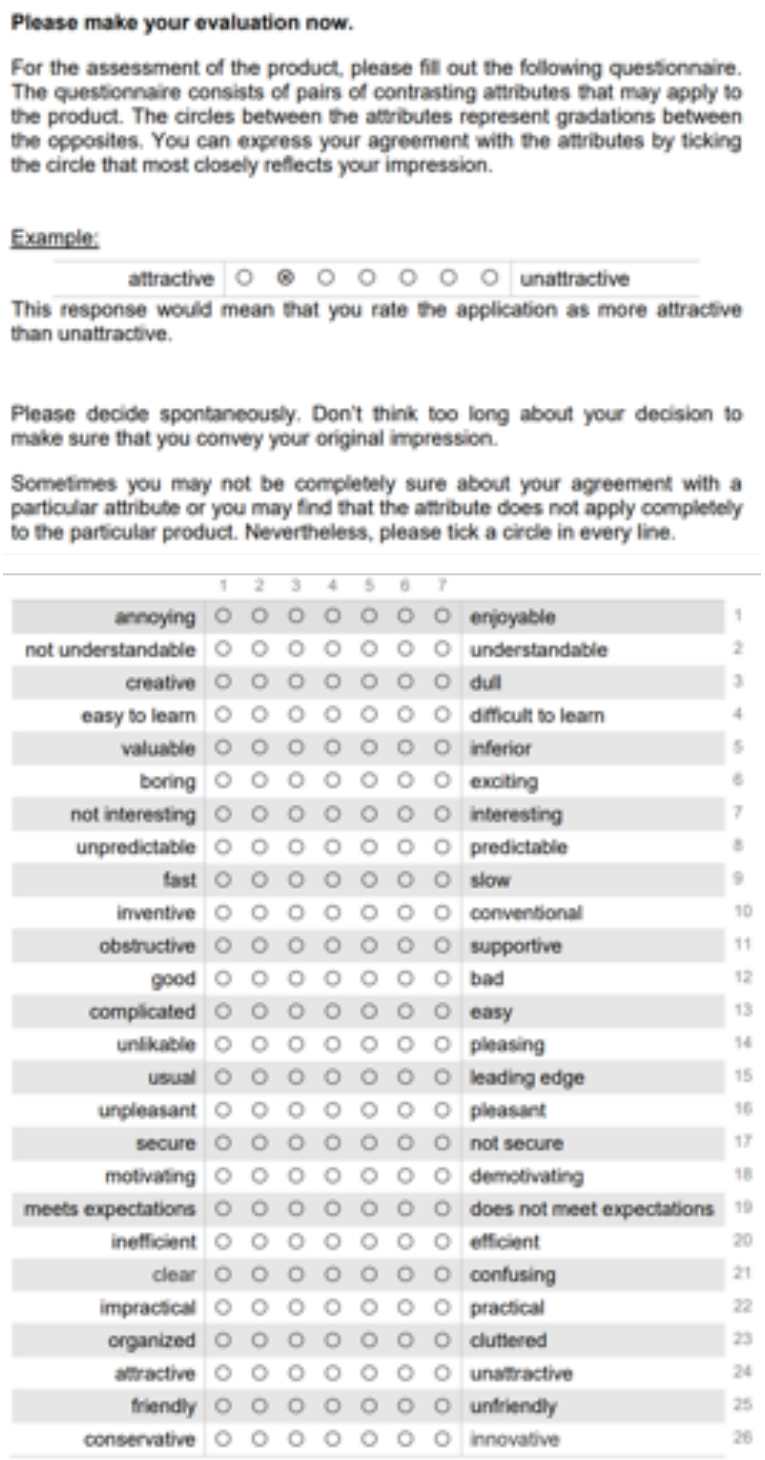


Figure G.11: Right Dorsiflexor repetition from the Strength Exercises with obstacles.

H

User Experience Questionnaire (UEQ)



Please make your evaluation now.

For the assessment of the product, please fill out the following questionnaire. The questionnaire consists of pairs of contrasting attributes that may apply to the product. The circles between the attributes represent gradations between the opposites. You can express your agreement with the attributes by ticking the circle that most closely reflects your impression.

Example:

attractive	○	⊗	○	○	○	○	○ unattractive

This response would mean that you rate the application as more attractive than unattractive.

Please decide spontaneously. Don't think too long about your decision to make sure that you convey your original impression.

Sometimes you may not be completely sure about your agreement with a particular attribute or you may find that the attribute does not apply completely to the particular product. Nevertheless, please tick a circle in every line.

	1	2	3	4	5	6	7		
annoying	○	○	○	○	○	○	○	enjoyable	1
not understandable	○	○	○	○	○	○	○	understandable	2
creative	○	○	○	○	○	○	○	dull	3
easy to learn	○	○	○	○	○	○	○	difficult to learn	4
valuable	○	○	○	○	○	○	○	inferior	5
boring	○	○	○	○	○	○	○	exciting	6
not interesting	○	○	○	○	○	○	○	interesting	7
unpredictable	○	○	○	○	○	○	○	predictable	8
fast	○	○	○	○	○	○	○	slow	9
inventive	○	○	○	○	○	○	○	conventional	10
obstructive	○	○	○	○	○	○	○	supportive	11
good	○	○	○	○	○	○	○	bad	12
complicated	○	○	○	○	○	○	○	easy	13
unlikable	○	○	○	○	○	○	○	pleasing	14
usual	○	○	○	○	○	○	○	leading edge	15
unpleasant	○	○	○	○	○	○	○	pleasant	16
secure	○	○	○	○	○	○	○	not secure	17
motivating	○	○	○	○	○	○	○	demotivating	18
meets expectations	○	○	○	○	○	○	○	does not meet expectations	19
inefficient	○	○	○	○	○	○	○	efficient	20
clear	○	○	○	○	○	○	○	confusing	21
impractical	○	○	○	○	○	○	○	practical	22
organized	○	○	○	○	○	○	○	cluttered	23
attractive	○	○	○	○	○	○	○	unattractive	24
friendly	○	○	○	○	○	○	○	unfriendly	25
conservative	○	○	○	○	○	○	○	innovative	26

Figure H.1: English Version of the User Experience Questionnaire (UEQ).

Por favor dê-nos a sua opinião.

A fim de avaliar o produto, por favor preencha o seguinte questionário. É constituído por pares de opostos relativos às propriedades que o produto possa ter. As graduações entre os opostos são representadas por círculos. Ao marcar um dos círculos, você pode expressar sua opinião sobre um conceito.

Exemplo:

Atraente	○	●	○	○	○	○	○	**Feio**

Esta resposta significa que avalia o produto mais **atraente** do que **feio**.

Marque a sua resposta da forma mais espontânea possível. É importante que não pense demasiado na resposta porque a sua avaliação imediata é que é importante.

Por favor, assinale sempre uma resposta, mesmo que não tenha certezas sobre um par de termos ou que os termos não se enquadrem com o produto.

Por favor, marque apenas um círculo por linha.

	1	2	3	4	5	6	7	
Desagradável	○	○	○	○	○	○	○	Agradável
Incompreensível	○	○	○	○	○	○	○	Compreensível
Criativo	○	○	○	○	○	○	○	Sem criatividade
De Fácil aprendizagem	○	○	○	○	○	○	○	De difícil aprendizagem
Valioso	○	○	○	○	○	○	○	Sem valor
Aborrecido	○	○	○	○	○	○	○	Excitante
Desinteressante	○	○	○	○	○	○	○	Interessante
Imprevisível	○	○	○	○	○	○	○	Previsível
Rápido	○	○	○	○	○	○	○	Lento
Original	○	○	○	○	○	○	○	Convencional
Obstrutivo	○	○	○	○	○	○	○	Condutor
Bom	○	○	○	○	○	○	○	Mau
Complicado	○	○	○	○	○	○	○	Fácil
Desinteressante	○	○	○	○	○	○	○	Atrativo
Comum	○	○	○	○	○	○	○	Vanguardista
Incómodo	○	○	○	○	○	○	○	Cómodo
Seguro	○	○	○	○	○	○	○	Inseguro
Motivante	○	○	○	○	○	○	○	Desmotivante
Atende as expectativas	○	○	○	○	○	○	○	Não atende as expectativas
Ineficiente	○	○	○	○	○	○	○	Eficiente
Evidente	○	○	○	○	○	○	○	Confuso
Impraticável	○	○	○	○	○	○	○	Prático
Organizado	○	○	○	○	○	○	○	Desorganizado
Atraente	○	○	○	○	○	○	○	Feio
Simpático	○	○	○	○	○	○	○	Antipático
Conservador	○	○	○	○	○	○	○	Inovador

Figure H.2: Portuguese Version of the User Experience Questionnaire (UEQ).

Nr	Item	1	2	3	4	5	6	7	Scale
1	annoying/enjoyable	0	0	0	1	0	3	1	Attractiveness
2	not understandable/understandable	0	0	0	0	0	3	2	Perspicuity
3	dull/creative	0	0	0	0	1	4	0	Novelty
4	difficult to learn/easy to learn	1	0	0	0	1	0	3	Perspicuity
5	inferior/valuable	1	0	0	0	1	3	0	Stimulation
6	boring/exciting	0	0	0	1	2	1	1	Stimulation
7	not interesting/interesting	0	0	0	0	1	3	1	Stimulation
8	unpredictable/predictable	0	0	1	0	1	2	1	Dependability
9	slow/fast	0	0	0	2	1	1	1	Efficiency
10	conventional/inventive	0	0	0	1	1	2	1	Novelty
11	obstructive/supportive	0	0	0	0	2	2	1	Dependability
12	bad/good	0	0	0	0	0	0	5	Attractiveness
13	complicated/easy	0	0	0	0	1	2	2	Perspicuity
14	unlikable/pleasing	0	0	0	0	1	2	2	Attractiveness
15	usual/leading edge	1	0	0	0	3	0	1	Novelty
16	unpleasant/pleasant	0	0	0	0	1	3	1	Attractiveness
17	not secure/secure	0	0	0	0	0	1	4	Dependability
18	demotivating/motivating	0	0	0	0	1	2	2	Stimulation
19	does not meet expectations/meets expectations	0	0	0	1	0	2	2	Dependability
20	inefficient/efficient	0	0	0	0	1	2	2	Efficiency
21	confusing/clear	0	0	0	0	0	2	3	Perspicuity
22	impractical/practical	0	0	0	0	1	2	2	Efficiency
23	cluttered/organized	0	0	0	0	0	1	4	Efficiency
24	unattractive/attractive	0	0	0	0	1	3	1	Attractiveness
25	unfriendly/friendly	0	0	0	0	0	1	4	Attractiveness
26	conservative/innovative	0	0	0	0	2	3	0	Novelty

Table H.1: Table with the Questionnaire's Distribution of Answers per item.

UEQ Scales (Mean and Variance)		
Attractiveness	2,300	0,30
Perspicuity	2,150	0,89
Efficiency	2,100	0,64
Dependability	2,000	0,72
Stimulation	1,600	0,36
Novelty	1,400	0,71

Table H.2: Questionnaire's UEQ Scales(Mean and Variance).

Nr	Item	1	2	3	4	5	6	7	Scale
1	annoying/enjoyable	0	0	1	0	2	10	4	Attractiveness
2	not understandable/understandable	0	0	0	0	3	10	4	Perspicuity
3	dull/creative	0	0	1	0	1	8	7	Novelty
4	difficult to learn/easy to learn	0	0	0	0	0	6	11	Perspicuity
5	inferior/valuable	0	0	0	0	1	8	8	Stimulation
6	boring/exciting	0	0	0	0	7	7	3	Stimulation
7	not interesting/interesting	0	0	0	0	3	9	5	Stimulation
8	unpredictable/predictable	0	0	0	3	6	5	3	Dependability
9	slow/fast	0	0	0	3	4	5	5	Efficiency
10	conventional/inventive	1	0	1	0	1	10	4	Novelty
11	obstructive/supportive	0	0	0	0	2	8	7	Dependability
12	bad/good	0	0	0	0	0	4	13	Attractiveness
13	complicated/easy	0	0	0	4	4	5	4	Perspicuity
14	unlikable/pleasing	0	0	0	0	2	11	4	Attractiveness
15	usual/leading edge	0	0	1	2	3	7	4	Novelty
16	unpleasant/pleasant	0	0	0	0	1	9	7	Attractiveness
17	not secure/secure	0	0	0	2	2	2	11	Dependability
18	demotivating/motivating	0	0	0	0	0	6	11	Stimulation
19	does not meet expectations/meets expectations	1	0	0	0	1	8	7	Dependability
20	inefficient/efficient	0	0	0	0	1	12	4	Efficiency
21	confusing/clear	0	0	0	1	4	9	3	Perspicuity
22	impractical/practical	0	0	0	0	2	13	2	Efficiency
23	cluttered/organized	0	0	0	1	2	7	7	Efficiency
24	unattractive/attractive	0	0	0	0	1	9	7	Attractiveness
25	unfriendly/friendly	0	0	0	0	0	9	8	Attractiveness
26	conservative/innovative	0	0	1	0	0	12	4	Novelty

Table H.3: Table with the Application's Distribution of Answers per item.

UEQ Scales (Mean and Variance)		
Attractiveness	2,333	0,15
Perspicuity	2,015	0,23
Efficiency	2,015	0,16
Dependability	2,029	0,55
Stimulation	2,235	0,25
Novelty	1,897	0,70

Table H.4: Application's UEQ Scales(Mean and Variance).