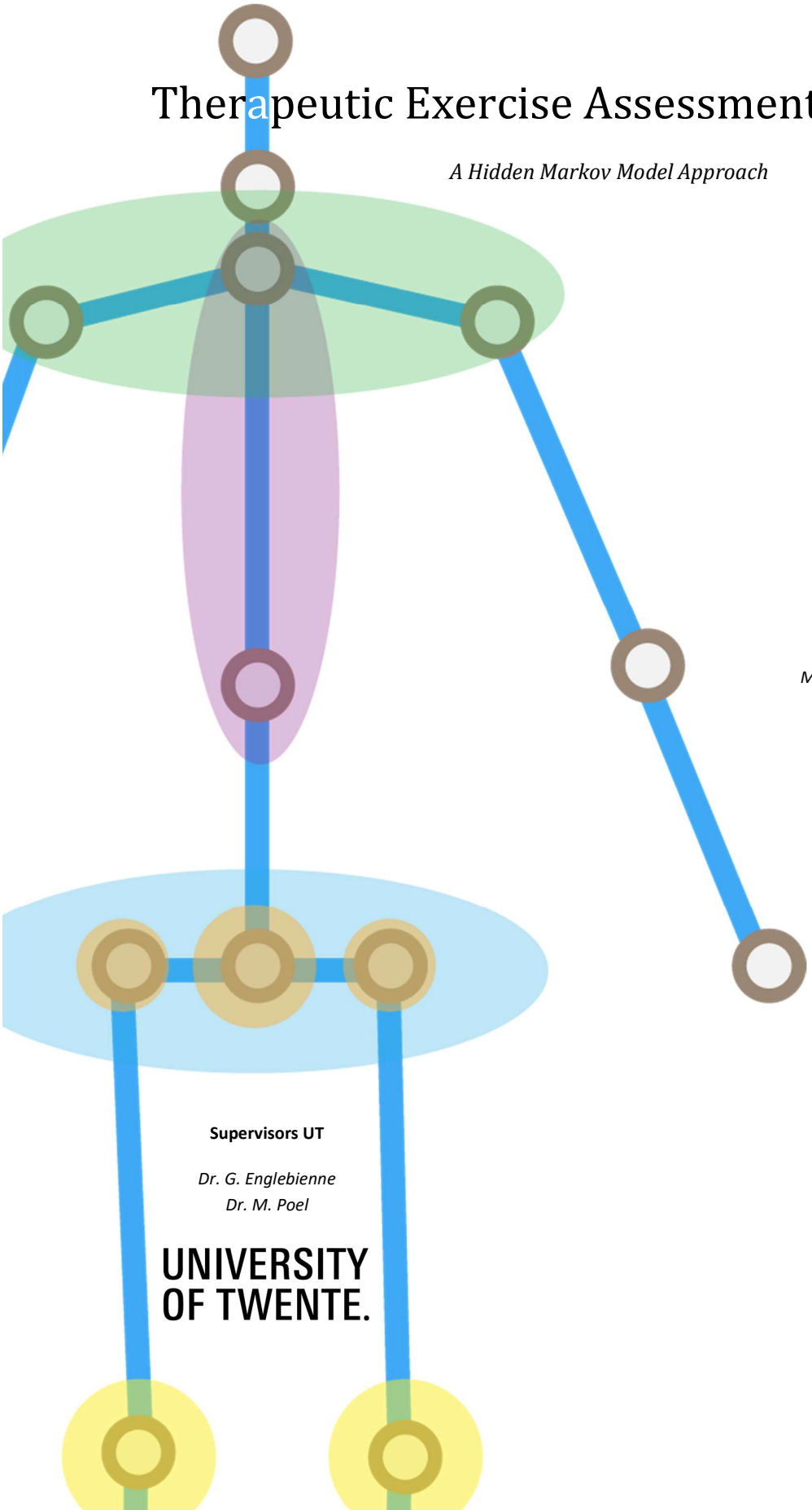
Therapeutic Exercise Assessment Automation

A Hidden Markov Model Approach



Jan Kleine Deters
M.Sc. Thesis
January 2018

Human Media Interaction
Faculty of Electrical Engineering,
Mathematics and Computer Science
University of Twente

Supervisors UT

Dr. G. Englebienne
Dr. M. Poel

UNIVERSITY OF TWENTE.

Supervisor UDLA

Dr. Y. Rybarczyk

UNIVERSIDAD DE LAS AMÉRICAS

Acknowledgements

The sunlight strikes my window every day around seven in the morning. Here in Quito the capital of Ecuador the foundation of this work is shaped. Not only the welcoming sun but the lovely people around me made me feel at comfort and gave me the confidence in adding value to the project I was working on. The collaboration with Dr. Y. Rybarczyk (head of project, Universidad de las Américas) was very pleasant as I felt as a bold discussion companion where I've gained a lot of insights in conducting scientific research and simultaneously my own strengths and weaknesses. The freedom to make mistakes pointed me on the effectiveness of 'doing' things with a knowledge gab that over time will be filled as practice makes perfect. As I know nothing is perfect, I can say with confidence that the biggest personal achievement has been the infiltration in this world of data science related fields and learning sufficient to provide the essentials for the project. This project being the development of a tele-rehabilitation platform, in short: a web solution to perform therapeutic exercises at home with automated feedback. Furthermore, I would like to thank the therapists involved in the project, Arián Aladro Gonzalo and Danilo Esparza, as they helped shape understanding of the whole rehabilitation process and did a lot of practical work. My supervisor at the University of Twente, Dr. G. Englebienne, has been very supportive during the entire process and allowed me to roam freely to discover the paths leading to therapeutic exercise assessment automation. In addition, Dr. R. Zalakeviciute embraced my practical machine learning skills and provided the opportunity to apply this in a different field as an extracurricular work. This collaboration increased my understanding and made me feel acknowledged as a starting researcher. This era of graduating made me realize I could prehab's continue in the world of research as I felt appreciated in making the first baby steps. Finally, I want to thank my family and friends, wherever they might be, for supporting me mentally, financially and emotionally.

Trabaja, juega y viva

Quito & Enschede, 2017

Abstract

In all kinds of research fields, the drive for automation seems to be rooted in the mindset of various stakeholders. Likewise, this is the case in clinical areas where specialist's time is valuable and should be directed to patients who need it the most. In the case of physical rehabilitation training, supporting patients (automatically) who can perform therapeutic exercises without physical support can save valuable time (for the therapist and patient) and improve the overall rehabilitation. The therapists should however still be in charge of the rehabilitation program and thus needs to receive a qualitative overview of their patient's exercise executions. Cost savings and flexible planning (ultimately progress estimation) arise when patients can perform exercises at home and receive therapeutic valid feedback. The goal in this work is to deliver the first constituent, that of automatically generating a valid therapeutic assessment.

In this work, a novel method in automatically extracting the quality in therapeutic exercises will be pleaded. The process in finding the right approach evolved from a literature review to an effort of iteratively crafting models and classification measures in an experimental setup (where data is gathered of the pursued exercises). This led to the construction of an assessment template that is as comprehensive as possible. Finding the key-elements in covering the assessment is further breed by a constant dialogue with various physiotherapists as they know, intrinsically, the art of 'seeing' quality and know how to approach the patients in a favourable way. Using data mining, time series analyses and machine learning in a hybrid fashion, the resulting methodology can be described as simultaneously a data and expert knowledge driven approach.

For the action/movement representation a 3D motion capture camera is used that provided additional processing capabilities, representing detected humans in a skeletonized fashion. From this bare human representation that purely indicates the spatial orientation of pre-defined joints, a personal and orientational invariant feature set is created that connects to the movement descriptions as practiced by therapists. With this invariant feature set different algorithms such as Dynamic Time Warping and Hidden Markov Models are trained on distinguishing good executions from bad ones, the type of error (also referred to as compensation) that occurred within an execution. Eventually the classifications of these models aid in detection of coordination errors.

As the assessment is not a standalone solution, a proposal on the integration of the suggested methodology can be contemplated in the final part of this thesis. In addition, developments within the umbrella project that will need to be accomplished due to this proposed method are some of the supplementary deliverables. A blueprint based on a Hidden Markov Model approach will provide insight on Range of a movement, compensatory behaviour, rigidness or instability, smoothness, pace, and coordination. This blueprint is the starting point for new research that can translate this pile of figures into comprehensible, valid and beneficial therapeutic feedback.

Keywords

Tele-Rehabilitation, Assessment Automation, Time Series Analyses, Therapeutic Exercises, Dynamic Time Warping, Hidden Markov Models, Human Invariant Motion Features, Data mining, Machine Learning, Therapeutic Feedback Generation

Content

List of Abbreviations

Well-posed problems

1. A solution exists
2. The solution is unique
3. The solution's behaviour changes continuously with the initial conditions.

Jacques Hadamard (1902)

1 Introduction

Project ePHoRt started in 2016 with the goal to create a web-based platform in tele-rehabilitation. This platform aims to rehabilitate patients recovering from hip surgery until they can walk again without any means. Based in Quito, Ecuador it is critical to consider the social and economic situation of the patients. Finding a low-cost solution is therefore the objective. In developing the platform, patients would gain several advantages over conventional physiotherapy (not that it will replace, rather complement). This includes the ability for patients to recover at home, thus reaching out to those who are not able to travel due to the remoteness of their habitat or mobility problems. The platform is however, by no means a replacement of conventional therapy but aims to be an addition to regular rehabilitation programs. At the same time, the cost of trainings sessions will be reduced, training can be personalized and remotely monitored. Training can also be executed more frequently and at preferred hours. In an earlier stage of the project, the basic architecture of the platform is designed, and an experiment preformed to select the most suitable technology to preform recognition tasks. These recognition tasks need to identify, specified exercises (by therapists), and difficulties or deviation in the patient's execution that are the results of incorrect execution of the therapeutic exercises. A vision based 3D-motion capture devise (Microsoft's Kinect) is suggested as the best fitting solution in this task. At a price of approximately 100 dollars is showed to be a valuable low-cost solution in comparison with high fidelity devises such as Vicon [1] or OptiTrack [2]. In addition, the preparation that is needed whiles using wearables creates additional interactions that distracts from the intended interaction and as such is undesirable. The next step in the development of the platform is to construct a robust method to assess the quality of executions of the therapeutic exercises. This will be the central focus point within the thesis as will be pointed out shortly.

1.1 The ePHoRt Platform

The platform aims to include interfaces for all stakeholders that are personally involved in the rehabilitation program (Figure 1.1). Interfaces in this case enables the medics to include information to the platform about any patient, such as: which hip, severity (how many stiches e.g.), movement restrictions, estimated rehabilitation and basic personal information. The therapists will be able to create a training, follow the progress of the patient and is able to autonomously gather data and create assessment tools. The patient will receive on the other hand, feedback that will aid in bettering the execution of the following exercises. In this thesis, the symbiosis of accuracy in assessment and implementation arises as the assessment tool should be understandable on an intuitively level.

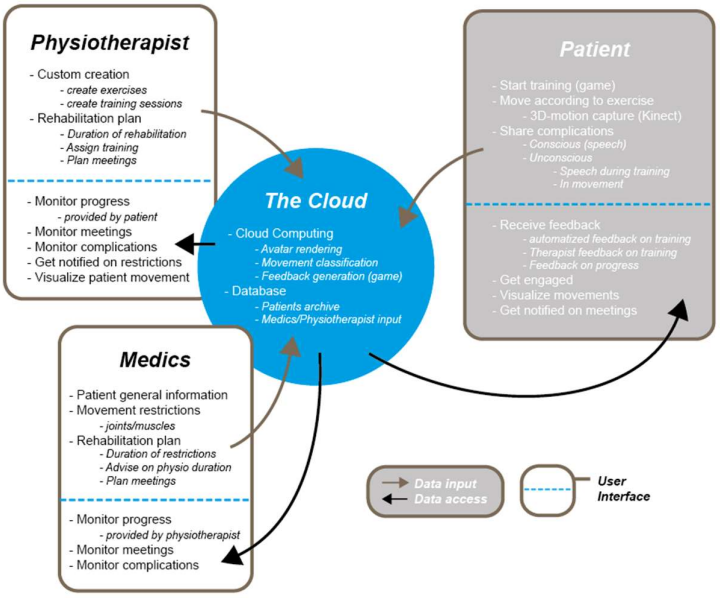


Figure 1.1 Overview of the architecture of the ePHoRt Platform

1.2 Research topic

How to assess the quality within a movement is the topic researched in this thesis. This ranges from investigating how data can be gathered and represented (pose extraction, feature creation) to efficiently be introduced into the assessment, reviewing algorithms that can aid in the process of assessing this data and more practically the recording and coding of exercises to carefully crafted categories. An orientation on similar existing projects sketches a basic idea of do's and don'ts as well as the undiscovered ***gabs*** in this research field. In collaboration with affiliated physiotherapist whom are also researches, understanding the mind of the human assessment is grasped and realized into schemes that can recreate the virtual assessment closer to that of the therapist. Conducting experiments is the foundation on which any claim/proposal will be derived from. These experiments are thoroughly designed to answer some of the research questions that are presented by now. The intention is to start with a pre-recorded set of data to build preliminary detection models and, gradually, increase the database with patient's data, to enhance the accuracy and generalization of the classification models. These recordings that will be used to train different models are likewise a result of this work.

Research questions

How can the quality within therapeutic physical exercises be assessed (automatically)?

So that it is easy to use by non-data specialists, provides quick insight, and provide therapeutic valid feedback.

What are the existing approaches (time series techniques)?
How are these integrated/used in tele-rehabilitation platforms?
How do therapist preform an assessment and describe movement?
How can the retrieved exercise be translatable to therapist/user feedback?
How can the assessment be integrated into the proposed platform?

1.2.1 Claim

The assessment of quality within movement is a new way of using gesture in interaction. Where there are tons of examples of gestures being classified as type 1 or of type 2, quality here is a scaling factor that should be represented as such. The difficulty in this area is that some exercises do not change in quality values when there are input differences. One example could be that during lifting your leg, the extreme pose could be hold for 1 or 2 seconds and both executions are equally correct. Therefore, getting into the mind of a therapist is essential in the prosses during this thesis. Shaping concepts, based on their beliefs and desires (as they also know what is best for the patient) will enhance the assessment. Mimicking the expert user in their decision process and getting to their tacit knowledge is even so more important, than creating accurate classification models. The accuracy however can he held as an indicator of the quality of extraction of the tacit knowledge. As when this underlying process, which is the go to method to rehabilitate, is uncovered, the only question left is a matter of engineering. This mimicking will ooze out through the whole motion to interaction process as visualized in Figure 1.2.

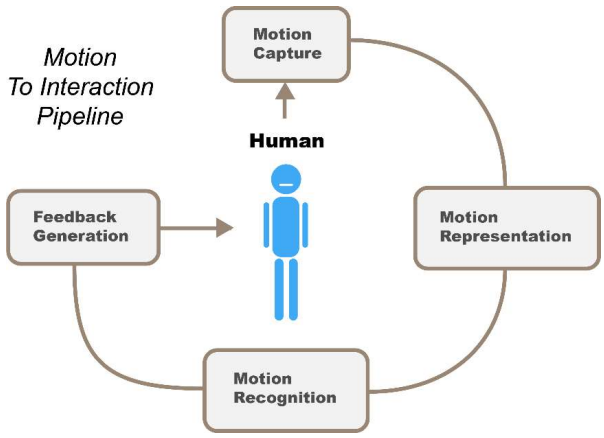


Figure 1.2 Motion to interaction, stages of data processing

1.2.2 Purpose

Creating an assessment method based on expert knowledge can be referred to as a classical example of a design problem that could be approached as top-down or bottom-up. Sophisticated algorithms such as deep representation learning show that "*it is possible to train neurons to be selective for high-level concepts using entirely unlabelled data"* [3]. This means that the thought process could be extracted if sufficient data is available. In the case of this platform, the data is yet non-existing and using existing databases limits the freedom of both the therapists and us (the platform developers). Therefore, a bottom-up approach will act in this work. This bottom-up approach could impact the scientific community by uncovering what meaningful movement assessment means and how a representation of the quality can be optimized. A robust method that could be easily adopted by researchers whom are involved in interactional areas would be a valuable achievement. With this, human motion based interaction could profit vastly as the methodology will inherently be non-use-case (hip exercises) specific.

The social impact this project hopes to induce is to improve the quality of rehabilitation by making this platform widely available, affordable and adaptive on personal preference and capabilities. Implementing the case study can provide insights in which focus points for which stakeholders are important and detect difficulties in communication, expectations and technicalities. It can frame the needed conditions to create tele-rehabilitation platform or comparable ambient assisting living applications.

1.2.3 Goals

There are a couple of goals that will be brought forward in this thesis. First, the main goal is to create a method that will allow for a correct assessment of later specified therapeutic exercises. Doing so, the method should provide an understandable classification values of relevant errors. This means that localization of an error is an important goal and in addition an 'easy' translation from classification results to human/therapist jargon will be focal points. This is of course a consequence of the implementation as desired. Constructing experiments and classification models in such ways that misclassifications could be analytically tackled will steer to the goal of understanding the data obtained from therapeutic exercises.

1.3 Thesis organization

This work consists roughly out of three distinct parts, which will be shorty described here. The related work section will explore the developments in the field of tele-rehabilitation, insights in therapeutic methodology (assessing quality within exercises), the process of pose extraction from sensorial data and implementation strategies of smart algorithms for motion recognition. Requirements on user, system, classification methods and experimental protocol will arise from this first part.

With the requirements, the second part tries to incorporate these and answers questions on implementation strategies, limitations and gabs in the exploration area. 4 different experiments are carried out to shape understanding in a legitimate approach of performing quality assessment to enable suitable therapeutic exercise feedback.

The manner of assessment is presented as a proposal for implementation into the ePHoRt platform. The proposed methodology influences in terms the structure of the platform. These structural changes (additional modules) are also part of this final constituent. The modules that are proposed aim to deliver user freedom, especially for the therapist as the platform should not have to provide a data specialist to be operational. The development process (method) and new insights that emerged during the quest in finding an optimal assessment methodology concludes the thesis in a conclusion and discussion for future work.

Related Work

2 Developments in Tele-Rehabilitation

As digitalization and automatization progresses, a new field of tele-medicine is explored. The goal of Tele-rehabilitation is mostly to aid in the improvement of patient's rehabilitation by means of availability and flexible use (at home). The research community is exploring possible solutions for sensing human motion, assessing motion, proving feedback and engaging the patients in the process all at ones. This, in combination with inter-user communication and the possibility for therapist to stay in control of the rehabilitation process, makes these platforms vastly complex. Commercial platforms are emerging such as *Valedo* by *Hocoma* or *The Biogaming platform* by *Biogaming*. These platforms make use of inertial sensors or visual based 3D motion capture sensors. Their focus seems to revolve around engaging the user by applying a trending technique called gamification. Established game platforms such as *Xbox* by *Microsoft* or *Wii* by *Nintendo* have already been creating gadgets that enables the integration of gesture based controls into their gaming environments during the past decades. Current research focusses on professional implementation strategies where different end users (patient, therapist, surgeons) needs are considered to shape concepts.

2.1 Interaction designs

Until this era of virtual assistants in rehabilitation most commonly physical interactions in gaming revolve around maintaining a healthy physical fitness. Apart of training physical health, cognitive [4] and social/emotional [5] heath are included in games as focus points in treatment of some kind. These factors can play a role as mental abilities can be affected after surgery and immobility can lead to social isolation. One example of rehabilitation treatment where the cognitive functioning is trained utilizes brain computer interfacing (BCI) [6]. The interactions embedded in the tele-rehabilitation platforms need to take into account user specific functioning regarding these different types of heath as well as cultural and demographical distribution of this population (e.g., elderly) [7] as sensorial functioning and engagement mechanisms can be influenced by these facts.

2.1.1 Training and motion capture

Commonly used exercises in proposed tele-rehabilitation platforms are based on a Task-orientated training (TAT). These tasks are real-life examples of practices in which the user may experience difficulties regarding the damaged functioning. With increasing difficulty of these tasks, the users recover to the point where they can function normally again, in daily life, without support. Different approaches are taking on sensing the movements of patients during computer assisted rehabilitation experiments. There are projects in which daily objects are used in combination with touch screens [8][9], wireless body sensor networks [9][10] (inertia sensors, e.g.) and most commonly practiced 3D visual based motion capture [2][11][12][13] techniques. Figure 2.1 provides an overview of the currently possible types of human motion capture for tele-rehabilitation purposes.

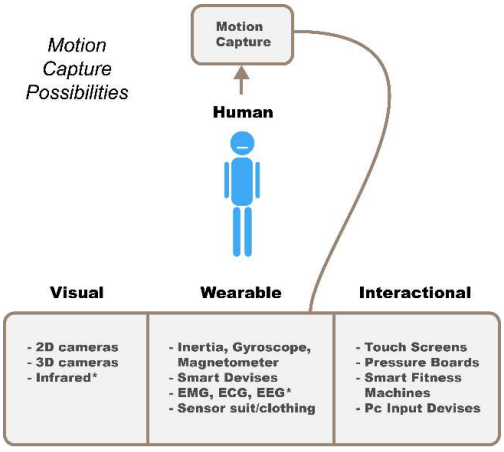


Figure 2.1 Overview of human motion capture possibilities

2.1.2 Interaction

Exploited interaction types in different tele-rehabilitation programs are (I) conventional therapy extended to the virtual where a personal coach is still present [14] , (II) fixed sets of exercises that need to be executed with automated quality recognition , (II) dynamically adapting virtual environments empowering engagement trough gamification [7] and (IV) Robot mediated with robots that are designed to interact in social situations [15]. Potential new interactions arise with the development of wearables where these devises not only could capture activity but generate feedback in a multi-sensorial fashion. Wearables such as exoskeletons are used to monitor in hand training [16] where the addition of tactile feedback (vibrations, applying external force) could lead to new interaction type; *"as if a coach would push you in the right posture"*. Most of these interactions are concerned with purely the quality recognition of the intended exercise. In addition, insight in mental and the physical condition, influenced by the execution of exercises, can aid in improving the rehabilitation process. Paralinguistic features obtained from voice or posture recognition could provide insight in one's mental/emotional status and could reveal the workload that is experienced. Sensors that measure physiological phenomena such as breathing [17], heart rate devices and measurement of oxygen saturation [18], could also provide insight in the workload that is experienced. These two additional measures can enable interactions to become more engaging and tailored to user specific situations. A guidance to use these qualitative and non-qualitative measures for interactive integration, such as in the gamified proposals, is to take into account the theory of flow (Figure 2.2) [19]. Here the mental status and performance is expressed as a relation between the user's skill level and the challenge of an execution. The optimal interaction is constructed as a cocktail of the areas where users are in arousal, flow, control and relaxation. In gamified environments, the user often competes against bots or other human beings to balance on skill level, creating goals and social meaning. These gaming elements could be used in tele-rehabilitation to connect fellow sufferers through completion and cooperation.

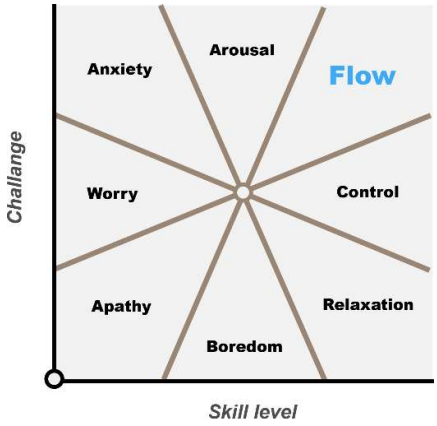


Figure 2.2 Implications (on mental state) of ratios between challenge and skill levels

2.2 System designs

The aim of several projects [10][14][15][20][21] is to create a web-based platform so that the user created content (exercises) can be stored, examined (automatically) and shared with therapist. On the other hand, not only the interaction for rehabilitation is shaped. Additionally the way that therapist can craft the exercises and training is a crucial part [2] in creating platforms that can be used in professional settings as partial replacement or addition to conventional physiotherapy. Low cost implementation [20] is focus for those who try to develop these systems intended for a large population and communities with less wealth. As tele-rehabilitation platforms use overlapping architectures, applications emerge that can aid in the platform development [22]. This versatile and integrated system for tele-rehabilitation (VISYTER) tries to aid in creating a solid basis to build on top on. Functions that are most commonly shared in the platform are data-base and communicational features between the different users.

2.2.1 Exercise Creation

In most research, the exercises that need to be executed, by the users, are initially recorded. New executions are compared with these recordings through the use of Machine learning algorithms [10], Answer Set Programming [2] and other algorithms that can cope with time-series data such as Dynamic Time Warping (DTW) [23]. Exercises are described by automatons [2], detecting the beginning and end of the exercise and discriminating idle movements from class specific executions. In some cases recordings only exist out of key-poses with interpolated paths [11] as reference exercise. Recording modules within platforms enable the therapists (Caregivers) to create exercises and user specific training. In addition to recording, exercise boundaries can be set by therapist [2].

These boundaries are deviations describing time durations or joint specific speed and range. The representation of the exercises takes the form of avatars [13], trajectories of the executed movements [9][24] and generally real-time video streaming of the user.

2.2.2 Research perspective

The last decade lots of initiatives in tele-rehabilitation has started. In Table 2.1 an overview is given on the basic types of rehabilitations. Most platforms only concern the patient or additionally the therapist that guides the rehabilitation process. Few works include other users than these two groups. Another noticeable fact is that the development of general use case scenarios is neglectable. It seems that specialized systems are more demanding or pave the road to general purpose systems as they are sole prove of concepts. In Table 2.1 V stands for Visual, W for Wearable and I for Interactional.

Table 2.1 Tele-Rehabilitation experimental developments

Year	Rehabilitation	Capture	Interaction	Users	Ref.
2005	Physical training	W	Extensive Sensor network	-	[17]
2008	After surgery	W	Sensor network, web-based, video/audio chat	Care takers	[10]
2010	Post-stroke	W+I	Pc, touch screen, sensor network	-	[8]
2011	Lower body	V	3D motion capture, screen, avatar	-	[13]
2012	Cognitive, Cancer patients	I	Exercise training, PC + input Devices	Care takers	[4]
2013	Post-stroke	I	Tablet, daily objects, Game	-	[9]
2013	Arm training	V	3D motion capture, voice command, Screen	-	[11]
2013	body scheme dysfunctions	V	3D motion capture, screen	-	[12]
2014	Fall prevention	V+W+I	Game, Web-based, TV + 3D motion capture, sensor network, pressure mat	Family, care takers	[7]
2014	Upper body, Cognitive	V	3D motion capture, screen	Care takers	[25]
2015	Autism	V+I	Robot (NOA), Web-based	-	[15]
2015	Hand training	W+I	Hand exoskeleton, Pressure sensitive ball	-	[16]
2015	Post-stroke	V+W	3D motion capture, screen, Oximeter (Oxygen saturation, pulse + inter beat value), Game-based	Care takers	[18]
2016	Arm training	W	Web-based, sensor network, PC screen	-	[20]
2016	Hand rheumatism	I	Exercise training, Web-based, Handhold pressure devise, Pc screen	-	[21]
2016	Chronic illness	V+W	Community interaction, supervised training, Video and audio communication, Web-based	Care takers, Fellow sufferers	[14]
2016	Pulmonary	W	Step-meter	-	[26]
2016	Post-stroke	V	3D motion capture, screen	Care takers	[2]
2017	Phantom Limb Pain	V+I	Mirror therapy, Tablet	Care takers	[27]
2017	Physical training	V	Video and audio, drawing on screen	Care takers	[24]

3 Exercises and Training

Rehabilitation revolves around enabling the patients to function normally in every day activity without discomfort or pain. This means that exercises created by therapists are designed to train ***strength, stability and flexibility*** of the affected muscles. After surgery, muscles can lose significant quality in these domains where hip muscle strength and leg press power after fast-track total hip arthroplasty can be reduced up to 58% [28]. Reduction in these focus points can lead to immobility, further muscles decrease and social isolation. Physiotherapists jargon and analyses is described here to understand the specialist's eye and shape concepts to represent capture data and show users their progress.

3.1 Surgery and Anatomy

Hip surgery is necessary when cartilage is damaged that used to enable smooth movement within the hip joint. After the replacement of parts of the hip joint, patients are expected to stay in the hospital up to three days [29]. The main affected areas (Figure 3.1, left), cartilage tissue on the socket and ball (Synovial joint) together with parts of the synovial joint are replaced with a prosthetic implant. The rehabilitation program normally starts the day after surgery and can last up to several months. There are two major risk after surgery, (i) blood clots due to inactivity and (ii) infections. Monitoring concerning infections could be done with the use of infrared cameras that can show increasing body temperatures that arise from these infections. During the surgery, different muscles are affected by the incisions that enable the reach of the joint. These muscles connect to the hip-bone, spine, upper and lower leg (Figure 3.1, right). Muscles that need to be mostly focused on during rehabilitation are leg and buttocks muscles.

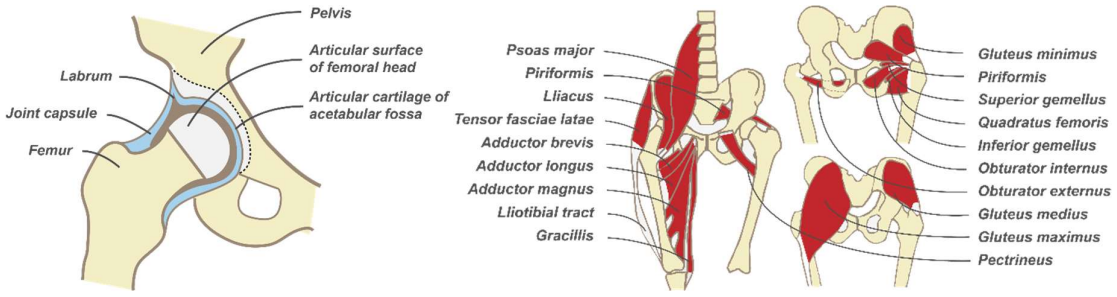


Figure 3.1 Left: The hip joint and affected parts before surgery, Right: The hip joint and its connected muscles

In the first stage after surgery, the healing tissue prohibits movements that involve strain on muscles and skin. As the post-surgery wounds are recovering, strain could cause wounds to take longer to heal or accidentally brake open. These movements are categorized as early stage harmful movements in the rehabilitation process. A category of movements that should not be performed up to a year after surgery include pivoting (rotate hip around spine without moving legs), twisting the leg, cross leg past the body midline and flexing the hip past 90 degrees. These movement can cause damaging and even dislocation of the implant. A second category of movement that is less harmful but merely an indicator of a lag of recovery are compensatory movements. These movements are characterized as a systematic deviation from a correct execution caused by unburdening unrecovered or weakened muscles.

3.2 Motion analysis

Human motion is expressed in reference to the posture as displayed in Figure 3.2. Movement can be described in three coordinate planes with respect to the subject. These planes intersect the subject's body at the spine where the transverse plane divides the body in upper (superior) and lower (inferior) part. Motion can be described in terms of the plane that they occur in. In the sagittal plane, flexion decreases the angle of 2 body parts and extension increases the angle. In the frontal plane abduction increases angles between body parts and adduction reduces it. In Figure 3.3 the terminology of hip movement is visually described. Motion analyses tools are available to track (marker based) angular changes in the different planes within video images (Kinovea, https://www.kinovea.org/).

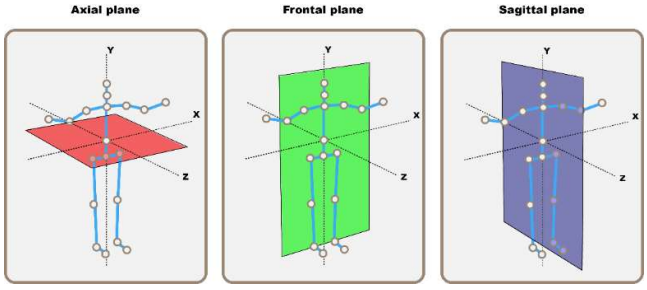


Figure 3.2 Planes of reference, movements are described from this personal perspective

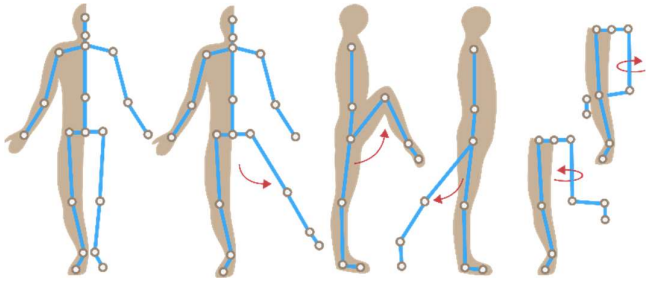


Figure 3.3 Reference body posture and Hip movement terminology from left to right: Abduction, Flexion, Extension, endorotation and exorotation

3.3 Qualitative analysis

Visual inspection is a common way to assess the quality of movement by trained specialists. There is a need for baseline measurements in physical abilities and physiological characterization of the patients to find suiting initial patient goals. These baseline test can involve measuring blood pressure, heart rate, pulse rate, temperature, joint range of motion (ROM) and Manual muscle strength (MMT) [30]. Furthermore, during the execution of an exercise articulation/coordination and compensations are measured to express the quality of the movement. The initial measurements describe the fitness of the patient and severity in defected functioning of the affected body parts. Heart rate can reveal the experienced workload. The ROM and MMT both can explain for reduced strength, whit ROM accounting for problems that can be related to flexibility as well (pain in flexion or dislocation). ROM can mostly be observed visually, whereas using MMT requires a tactile interface to measure force and or apply external forces to a specific part of the body. During the execution of a movement there are cues that can provide insight in the strength such as trembling, speed of movement and the time a certain pose can be hold. Besides these measurements that reveal the basic status of the patient, mental measurements play a key role during the rehabilitation as it can uncover the perceived pain. Pain can be a sign of inflammation and will influence the workload of the patient on a daily basis. Using the Pain Catastrophizing Scale (PCS), basic insight can be provided on the perception of pain by the patient utilizing questionnaires for example. The PCS focuses on the patient's tendency to communicate his/her pain, the magnification of their pain (future projection, e.g.) and a helplessness factor. The focus of an exercise is partially shaped by the force that needs to be applied during an exercise and the speed of the execution. The orientation of the patient (laying down, sitting, standing) influences the force application as in one orientation the muscles would not have to overcome gravity's pull in respect to another orientation. According to the stage (Strength, stability and flexibility) exercises are designed with a ratio that intersects the curve as in Figure 3.4.

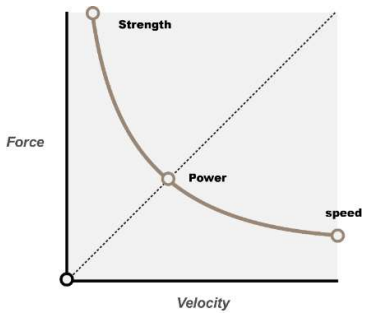


Figure 3.4 Force/Velocity ratios and related exercise focal point

3.3.1 Exercise Quality

During the execution of a rehabilitation exercise different types of quality can be observed, these can be seen in Table 3.1. Some of these types of executions have strict boundaries which means that a detection of such kind should immediately result in a feedback (harmful movements). Compensation movements describe an increased trajectory path or speed of any given body part excluding the target body part (hip flexion e.g.). If a compensation occurs, it can occur in a good or bad coordinated way. Good coordination is synchronized movement of this other body parts relatively to the target body part. A correct execution is an exercise in which none of the above types occur and where the desired range of motion is also reached. Except for the harmful types of execution, any execution type can take value out of a distributed area of multiple features/descriptors (described in chapter 4) with ambiguous overlapping regions where good shifts to bad. In these regions, the assessments between individual therapists could be inconsistent as per individual the mental mapping of these distributions can slightly differ due to training and/or practice.

Table 3.1 Types of exercise executions during rehabilitation

Stage	Types of executions	Consequence
5-30 days	Early stage harmful	Intervenes recovery wound
Up to 1 year	Harmful	Intervenes settlement implantation
Up to final rehabilitation months	Compensation	Weaker muscles (uninvolved)
Up to final rehabilitation months	Bad Coordination	Bad force distribution
Up to final rehabilitation months	Faulty	Slower or no rehabilitation
During total rehabilitation	Correct	Rehabilitation

3.3.2 Quality progress

During the whole rehabilitation program, the expected quality per execution increases with time. In the first stages of the rehabilitation the focus lies on reduction of all the unwanted types of execution. Simultaneously consistency should increase, meaning that the fluctuation between executions dissolve and every execution looks more like its predecessor. If there is a reasonable consistency and compensation/coordination stay within reasonable bounds (therapist's definition) the focus shifts on improving the range of motion. An overview of the rehabilitation definition and necessary steps to track the progress as stated here can be seen in Figure 3.5.

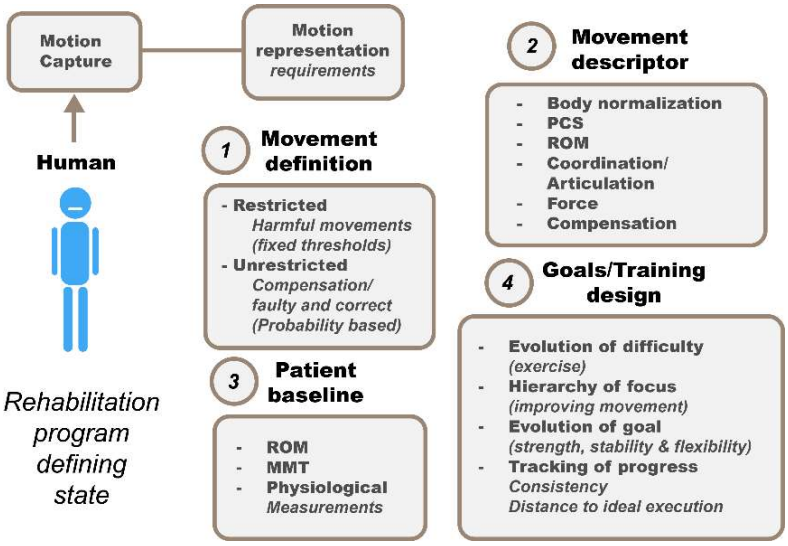


Figure 3.5 Rehabilitation program definition

4 Pose Extraction

There are several ways to record and extract human poses for interaction and monitoring purposes. A pose in this case is a 3D representation of the human body, defined by robust features. The robustness of these features enables the pose extraction technique to be generalized on a wide variety of human body types. Roughly three methods are used to record the ontological structure of a human being at a given timestamp. As mentioned in chapter 2 there are the non-invasive wearables, Interactive and vision based systems that will either capture 2D/3D images or both. In this chapter, the focus lies in providing an overview of the vision based approach to create feature representations that will enable the estimation of a pose where the influence of variance in lighting, position and body types are being diminished extensively.

4.1 Vision-based pose extraction

Most promising in robustly extracting human pose in a scene is by utilizing 3D capture techniques. This is because shapes are easier to discriminate as low contrast in colour takes no effect, and the structure of occlusions can be identified. 2D techniques however are generally more cost efficient, more widely available and most of all there is a great deal of historical data available to incorporate whiles training models. Different researchers have been creating databases revolving around specific human motion and are freely available for public use [31]. These mostly revolve around short gestures intended for interaction. The specialization on therapeutic exercises seems yet to be developed.

Most systems that create 3D representations use stereo cameras in combination with structured infra-red lighting. This structured light is shaped into a point grid or other recurrent patterns. As the light hits an object, it deforms. Then the differences in shifts per stereo image creates the knowledge of the distance (depth) of a point by appliance of triangulation. An alternative approach to structured light is a time-of-flight (ToF) measure that creates a distance map based on reflection time of plausible emission sources.

4.1.1 Articulated tracking

While extracting a human body pose in a specific articulation, articulated tracking describes algorithms that benefit from the knowledge of the human body in tracking. As body parts are rigid and joints have a limited degree of freedom (DoF), consecutive poses have dependencies. In this sense, an initialized body frame can be more efficiently recovered in a following frame as a probability estimation of the new positioning creates a vastly smaller search area and is thus computational beneficial with great use for real-time applications.

4.1.2 Training data

In both constrained (Skeletonization) and unconstrained pose extraction (Visual words) techniques, the first vase of the pipeline is recording human bodies performing different actions (in classification problems). In constrained models, the interest points need to be labelled to be able to infer these interest points within various poses. In unconstrained models, the recording of different poses leads to the shaping of interest points that can distinguishes one specific type of movement from another one. Training data can be complemented with artificial data [32] that can increase the coverage of the solution space and make classification less prone to overfitting.

4.1.3 Object segmentation

Object segmentation can be performed in varies ways such as utilizing planar segmentation algorithms [33]. Segmentation limits the scope of further analyses of the goal object, beneficial as analyses can focus on the properties of this goal object in various conditions. Thus, firstly objects need to be recognized as only then the segmentation process can be adequately executed to isolate regions of interest. Segmentation conditions in clustering regions rely on similarity, proximity and continuity [34]. These conditions can be analysed with colour, recurring patterns, edge/line detection, convolutional responses (morphological operations) and regional configurations of these elements embodied in descriptors such as Histogram of orientated gradients (HOG). In 3D images things get easier as multiple object do mostly not share a common depth as well as a common alignment in the camera to object path, however occasionally when objects interact this will occur. A depth image also allows for segmentation based on surface curvature and derived object volumetric and size. Object segmentation for interaction relying on sequential frames can utilize features that express movement such as Histograms of Optical Flow (HOF).

4.1.4 Object recognition

Regions of specific interest can be labelled as such, so that these regions can be trained on using all sorts of machine learning algorithms. These different algorithms will be discussed in a later stage. Areas in an obtained image can be scanned to match a learned distribution with a learned threshold to be classified as an area of interest. Recognition for dynamical objects, such as humans, is often split up parts (Deformable parts model, DPM) that will be recognized and can therefore immediately infer the articulation of the body [35][36] and provide a semantic description of the observed pose [37]. For interactive systems, such as tele-rehabilitation platforms the recognition trough movement is most straightforward where initially shifts in two consecutive depth images indicate movement at a specific distance of the camera. Segmentation can then isolate this depth region as a bounding box or detect the whole moving object trough continuity measurements of the surrounding depth pixels. The general pipeline for pose extraction can be seen in Figure 4.1.

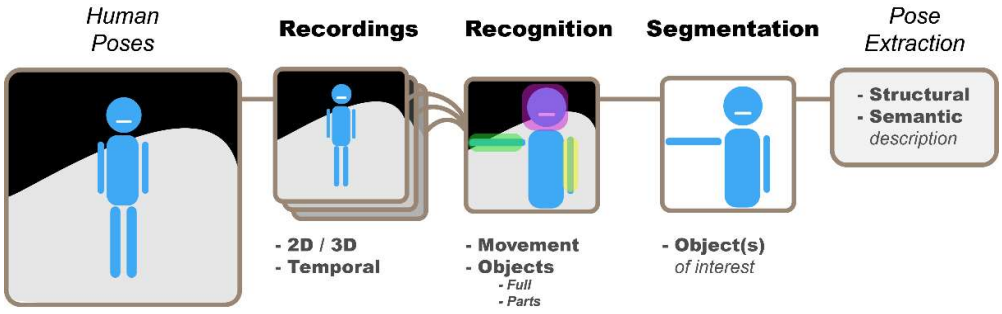


Figure 4.1 Pre-processing pipeline for human pose extraction

4.2 Featurization

A feature is an attribute of something, an attribute obtained by a stream of measurements/observations/values or assignments of values to the unknown distributions of these measurement values (labels). Featurization is used, as mentioned, to create robust representations of data that can be used to generalize with occurrence of various disturbances in a dataset. These disturbances are mostly influenced by environmental (lighting, objects, movements)/subject (shape, distance, orientation, etc.) changes but could also be caused by technical elements (calibration, recording speed, sensorial space dimensions (SSD)). A feature representation normally reduces the data that is observed into abstractions that can be used in classification, distribution learning for anomaly detection, and regression tasks. As example: a text can be represented as a sequence of distinct part of speech tags (PoST) that represent the amount and order (structure) of syntactical functions used. This can then be used to find language and user specific structures. In the case of human pose estimation, obtaining an abstracted form of the human body through Skeletonization (4.5) provides the relation between body joints. Different relationships within poses can, with the use of this feature representation, be distinguished from one another.

4.2.1 Vector Quantization

Vector quantization (VQ) is a way to compress data and can be used when efficient data storing is required. In classification problems, the implementation of vector quantization is merely the question of the mandatory accuracy. In the case of tele-rehabilitation platforms, the use of data compression can be valuable as the amount of data grows with every exercise executed by a patient. The advantage of this specific technique is that the compression is based on the distribution of the data, therefore high-density areas will be represented with more values after compression. Regions where data points get assigned the same values is bounded by a vector space (hyperplane) that result from the modelling of the probability density function of the data (figure 4.2). The centre of this vector space is the so-called prototype vector (the vector representation for the bounded area). VQ can be used for visual recognition problems to for example reduce the RGB colour space [38], when used as feature space. In supervised learning VQ can be used in finding class labelled prototypes. VQ is part of a branch of machine learning called competitive learning algorithms as D. Nova [39] stated and is related to algorithms such as Self-organizing (SOMs) maps and fuzzy c-means clustering.

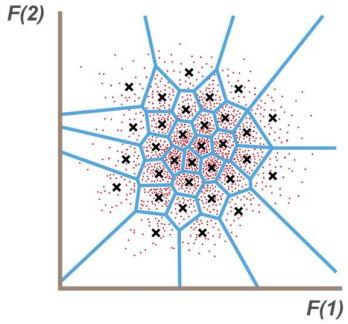


Figure 4.2 Example of Vector Quantization, in red the data points, blue the hyperplane intersections and in black the prototype vectors

4.2.2 Dimensionality Reduction

Dimensionality reduction is useful or can be useful for a couple of different reasons. The number of dimensions in classification firstly will extend the necessary learning time of a model and could influence processing time of the classification itself as well. More significant reasons to apply dimensionality reduction however has to do with improving the accuracy in classification and making the classification more robust (less sensitive to noise). Two conventional approaches are: (I) reducing the dimensionality whiles preserving the dimensions that describe the variation of the data in the best manner (Principal Component Analyses, PCA, Figure 4.3 left) and (II) preserving the dimensions that are most discriminative between pre-defined classes. A first step in dimensionality reduction is often the utilization of reduction method (II). This group of methods can be referred to as Discriminant Analyses (DA, Figure 4.3 right) and takes different forms in detail that can give slightly different results. Depending on the expected distribution functions between classes and the number of available samples, a specific type of DA can be selected in the reduction process such as mixture, quadratic, linear, flexible or regularized DA (MDA, QDA, LDA, FDA and RDA).

Another group of algorithms that can be used in reducing dimensionality are auto-encoders. Auto-encoders firstly compress data followed by a decompression of this compressed representation, using Neural networks (NN) in the process. These processes (en/de-coding) can be learned by using a distance function (loss function) applied to input and output data, and trying to minimize this loss in the learning process. Auto-encoders can be used as pre-processing step in creating generative models (*Variational auto*-encoders (VA)) for data generation or classification. In the case of temporal sequences, auto-encoders are composed of encoders and decoders that can capture this temporal structure. The temporal structure can be captured by long short-term memory (LSTM) networks, a type of recurrent neural network (RNN) that is able to learn dependencies over a longer time span.

Sparse coding is a type of dimensionality reduction where the input can be represented as linear combination of a set of vectors (input = $0.3*\text{vector}_1 + 0.6*\text{vector}_2$, e.g.) [40]. The goal is to have as least components (vectors) with non-zero values as optimization with an additional free parameter that resembles trade-off in this sparse representation to reconstructive power. Sparse coding is used in learning over-complete sets of vectors, where over-completeness means that the removal of a vector does not affect the reconstructive power.

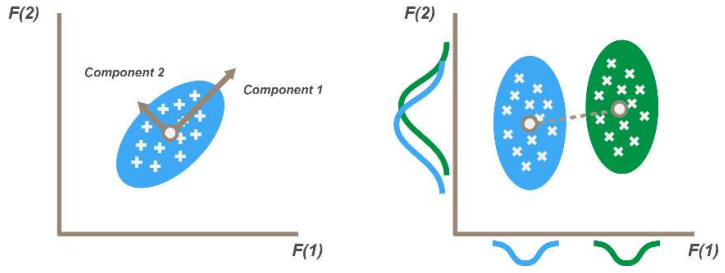


Figure 4.3 Example of PCA (left) and LDA (right)

4.2.3 Appearances

3D and 2D visual data can be represented in a multitude of appearances (Table 4.1). These representations can often already be used as valid features in classification. Here an overview is provided of the types of features that can be used for object and pose estimation. A typical approach is to create a histogram representation of parts of an image/3D scene, which eventually leads to a detection of some sort. Histogramming is creating a discreet representation of a signal and could be seen as a quantization method but is often not based on a density distribution of the data. There are however methods, filter banks like Mel-frequency cepstrum (MFC) that have unequally spaced vector spaces respectively to bin (discreet value of a histogram). Recent work involves creating the ideal filter banks (e.g. which areas take most variation) using deep neural networks [41]. Raw data such as pixels in a 2D image or points in a 3D cloud inherit multiple attributes that contextualizes these points. As humans classify based on these contextualizations, appearances that follow the same principle are created of images to feed into classification algorithms. Low-level appearances try to extract the contrast shifts (edges, corners) and orientations (Gabon, FFT) of continues elements. Histogram representations can represent a relational structure in various scales in terms of ontology and displacement (HOG, HOF). In 3D the structure can be described as curvature of a surface relative to a point of this surface. VFH is a representation that represents the mean curvature around each point in a point cloud and can therefore describe clouds invariant to camera roll-orientation and distance.

Table 4.1 Types of transformations for visual feature creation and complexity of the algorithm

Low-level	Medium-Level	Advanced-level
Edge detector	Viewpoint Feature Histogram (VFH)	Spatial temporal interest points (STIP)
Corner detection	Camera roll Histogram (CRH)	Scale-invariant feature transform (SIFT)
Gabor filter response	Histogram of Optical Flow	Deformable part model (DPM)
Fast Fourier transform	Histogram of oriented gradient (HOG)	

4.3 Pose based approach

Pose based approaches use the appearance of a human to draw conclusions by looking at a silhouette. This silhouette can be transformed into a convex hull representation that can be easier to examine on self-similarity and ratios within the silhouette. Obtaining a rigid silhouette after segmentation can be done by performing basic morphological operations (kernel convolutions) such as iterative dilation/erosion for gap filling [42]. Other Morphological operations can find parallel lines, skeletonize and find endpoints and line intersections in this skeletonized version. Here skeletonize refers to thinning the silhouette with keeping the ontological structure (holes) and orientation (Figure 4.4). Silhouettes are used in [43] to create motion-energy-images (MEI, where did movement occur) and motion-history-images (MHI, speed of movement related to spatial location) that capture shifts over time in these silhouettes. This MEI and MHI are then used in a template matching fashion where shape and intensity (amount of time a body part was in a certain position) are matched and variation is in this way spatially locatable. The curvature of the silhouette can also be used to distinguish for various body shapes.

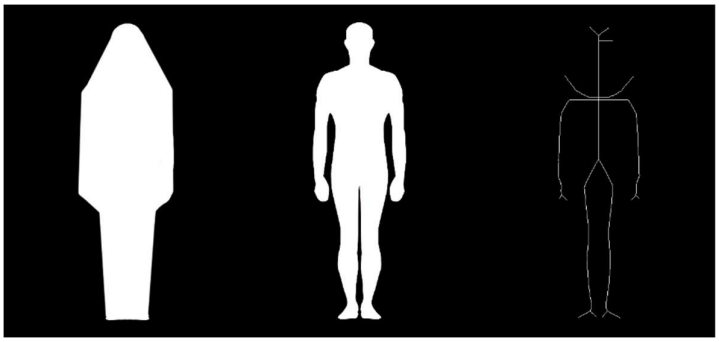


Figure 4.4 Left a convex hull representation, right a skeletonized representation of a binary image (middle)

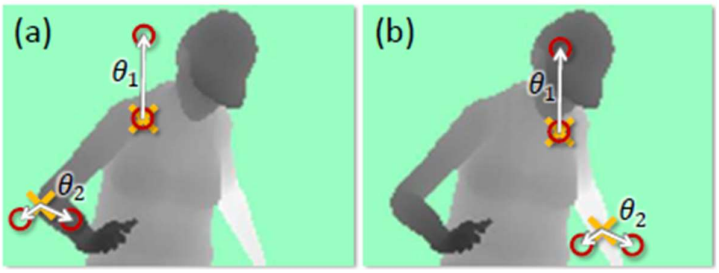


Figure 4.5 Features used in a Skeletonization algorithm that takes normalized offsets in an image to create a per pixel body part classification[1]

4.4 Skeletonization

Skeletonization essentially means that a skeleton is inferred from an image (colour, infrared, depth) if there are any humans in the scene. There is current research on not only inferring posture but also body shape [44], which can be useful in tracking progressions in muscle build up or loss of fat. The main problems of many approaches [45, Table 2] are that the processing memory requirements does not accommodate for a real-time application. Research has shown that gesture recognition vastly benefits from the prepossessing step resulting in Skeletonization [46], that is in comparison to the use of 'fast' low-level features. Skeletonization can roughly be split up in two steps, one in which an initialization takes place and a second step in which a tracking algorithm is involved. Here the first part will be discussed, later (4.3.10) the tracking will be discussed.

State of the art performance in real-time skeleton initialization is achieved by an algorithm implemented in *Microsoft Kinect's* platform [32]. Here real data is combined with artificially created data (of human bodies in different poses, in depth images) to train a random forest. A segmented depth image is the input of the training where the isolated and kept images are human representations. A per pixel low-level feature representation categorizes the body pixels into 31 body parts. A texture map with these defined body parts is wrapped onto the various training data as label. The feature representation (Figure 4.5) is an offset in depth between (I) two pixels, left and right to the target pixel with a distance between them that is normalized on the target pixel's depth and (II) an offset between the target pixel and a pixel above the target pixel (again normalized). Each tree in the forest is trained on a random subset of 2000 pixels per image. The output of every tree is a distribution of the likelihood that a pixel belongs to any body part. The outcome of all trees is averaged to create a classification. The position of the joints is then inferred by finding the mode of each region with using a mean shift based approach and a weighted Gaussian kernel. This pinpoints the joint position to be on the surface of the depth image. A learned offset between this point and the actual joint location is then used to advance the joint into the body.

[1] *Retrieved from [32]*

4.5 Anthropomorphic Constraints

Although there are differences between humans in the ratios of their respectively body parts, there are fundaments that provide estimated averages of human body ratios supported by for example natural ratios (golden ratio) expressed in geometrics or statistical data. These estimations are used in the arts, design and medicine to support users or depict visually appealing representations of a human. In the recent developments of pose extraction and motion estimation knowledge about the human body is introduced as constraints [47] to include statistical probability into the search problem.

The field of human ergonomics tries to make these measurements more explicit by capturing distributions of body measurements, Anthropometric Dimensional Data (ADD). This covers the probability of the occurrence of a length per observed body part. The correlation between individuals is not necessarily captured in the ADD. This correlation could however provide great advantage as a better anthropomorphic representation could be made if observations of a body part are ought to be highly certain. Anthropomorphic constraints can help in a more robust estimation of joint position and decrease the search space during joint tracking. Not only body lengths can be useful as a constrain, in addition the degrees of freedom per joint can be considered that will prohibit certain configurations of body orientations to be reproduced. Per subclass of movements the values can differ as for example a gymnastic could rotate their legs easily up to 30 ° [48], this range of individual flexibility can to some extend therefore be coupled to practiced sport and additionally age where an decrease of 6 degrees per decade in the range or 50 to 80 years old subject was observed [49] for ROM of the hip joint. As stated in [47] currently used models do not incorporate statistical anthropomorphic information. In this work, the auteurs add additional assumptions such as the presence of at least some parallel body segments to the image plane, this however constrains the method to only be useful if there is an initialization step (subject in base pose). In [50] not the DoF is used but rather the distance between any two skeletal points, also a symmetry is expected in the skeleton. Anthropomorphic constraints are often used in the skeletal retrieval in 2D representations as poses in these images can be highly ambiguous. Therefore, reducing the possible body orientations decrease errors and search time. In Tylor [51] the constraints are transferred from a 3D to a 2D projection with the skeletal size as input, according to the researchers with additional constraints the model is capable of creating unique solutions per analysed image. In [52] a normalization step, creates a reference ratio to the lower right leg and uses certain criteria for the expected ratios, these proportions are kept similar for every user which could decrease the potential accuracy for some individual cases. Figure 4.6 shows an overview of the updated motion to interaction pipeline as result of this section.

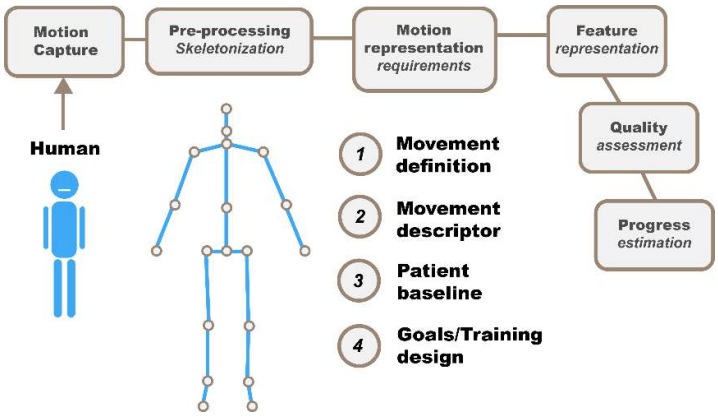


Figure 4.6 Motion to progress estimation

Intermezzo

Speech and language processing is a field that copes with similar problems as the field of motion recognition. In both cases, temporal patterns that are (highly) stochastic are translated into categories that represents distributions. These distributions can overlap in which case context does matter (e.g. what came before). This context is easy to see for language where we as humans sometimes also utilize the noisy channel model (NCM) to infer ones spoken words. In motion recognition tracking algorithms, can use learned constrains in human motion (limiting the prediction space). Likelihoods of transition from one pose to another as well as speed evolutions can be used in a similar NCM fashion where corrections in body positions within a sequence can be corrected to the most likely representation of the real motion.

Visual Words - A method developed to featurize multi-modal RGB-D data, based on an expansion of the bag-of-words method is the so called Bag of Visual and depth Words (BoVDW, Figure 4.7) [53]. The bag-of-words method is used in document classification and information retrieval. In this method grammar and order of the words do not play a role. Instead frequencies of 'tokens' (words) creates a histogram representation of the document. The BoVDW method utilizes depth descriptors that consider the normal vectors of the depth image with respect to the cameras position and rotation to its roll axis.

The extension to Bag of visual words means defining visual 'tokens'. First, images can be decomposed by means of spatial sampling or with the use of key-point extraction. The spatial samples can be clustered in N tokens trough clustering techniques that take a histogram representation as input (HOG, e.g.). Key-point tokenization can be performed by clustering key-points in the spatial domain, creating tokens that represents points as combinations of different types of key-points (corner + edge, e.g.).

In this method firstly Spatio-Temporal Interest Point (STIP) [54] are extracted. For each of these interest points, HOG, HOF, and HOG/HOF [55] combination descriptors are created over a 10×10 grid, for the RGB image. A VFHCRH (Viewpoint Feature Histogram, Camera Roll Histogram) descriptor is created for the Depth image.

VFH [33] creates a histogram of points in which the distribution of mean curvature around each point is represented. In this method, a whole point cloud is represented as a histogram containing 308 bins. VFH creates a histogram of the angles between surface normal for each point relative to its k nearest neighbours in a point cloud as suggested in the Fast Point Feature Histogram (FPFH) representation. VFH extents this with a viewpoint representation, creating a histogram of the angels of the surface normals with the 'general' viewpoint direction. This general viewpoint direction is the vector that is drawn from the camera to the centre of the surface of the segmented point cloud (object), making this representation scale invariant. (Segmentation with the use of depth image)

Using clustering technique over the total amount of descriptors trough k-means N words are extracted from video sequences. A sequence is represented as a spatio-temporal pyramid that introduces geometrical and temporal information. A 92 bins CRH is computed for encoding 6 DoF information of a point cloud.

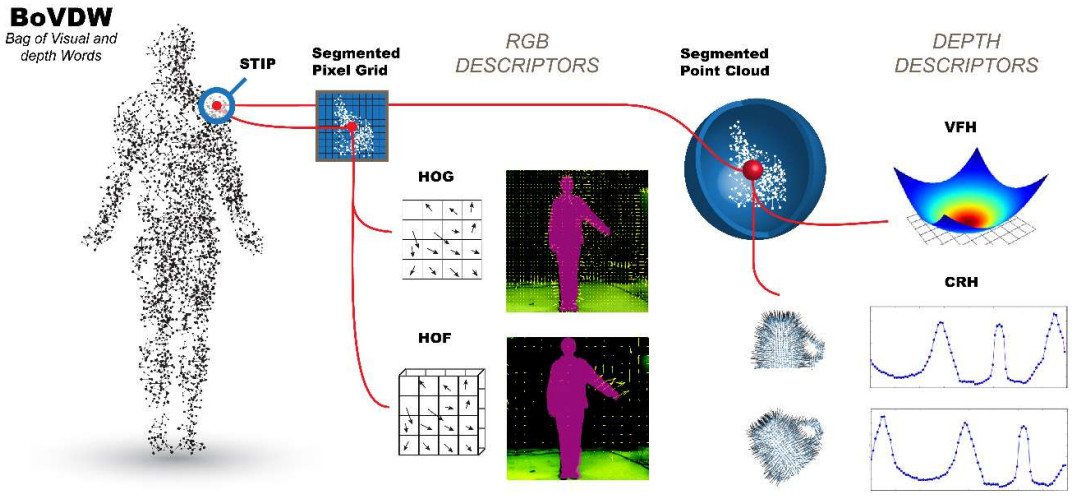


Figure 4.7 Bag of visual and depth words feature representation, combining HOG and HOF in 2D with VFH and CRH in 3D

5 Motion Recognition

Human motion analysis and classification are particularly challenging due to the intra and inter-individual variabilities in the execution of a same movement. Time duration, which is the main variance within human movement, can be coped through several time-series algorithms. When we talk about models that can predict or classify time series, there are a couple of distinctive categories that can be discussed. As stated in a review paper [56] there are various fields where time series analyses contribute in improving professions such as: Smart Surveillance, Behavioural Biometrics, Gesture and Posture Recognition and Analysis, Robotics, Medical, Sports and Exercise and Art and Entertainment.

Time series can be evaluated in the frequency or the time domain. One example where temporal signals are evaluated in the frequency domain, a currently popular application that enables you to find the title of the song you're listening to. This application (Shazam) creates a histogram and finds the most consistent match in a database that consist out of tons of histogram representations of songs. In this case analysing the time series in the frequency domain is useful as the signal itself is not stochastic in the time domain. This technique however is less useful when it comes to short-term meaningful motion recognition as a smaller 'listening' window will give poor frequency resolution (Fourier transform, uncertainty principle). Thus, in this section the techniques that are used to predict and classify time series in the time domain will be discussed. There are different families of algorithms that can deal with time-series data. Here the most common ones (Figure 5.1) will be discussed and evaluated on their applicability. In addition, data gathering, and data clipping will be discussed as these are required to be handled adequately before classification models can be implemented. The section will be concluded with an overview of research that implements variations of these different models.

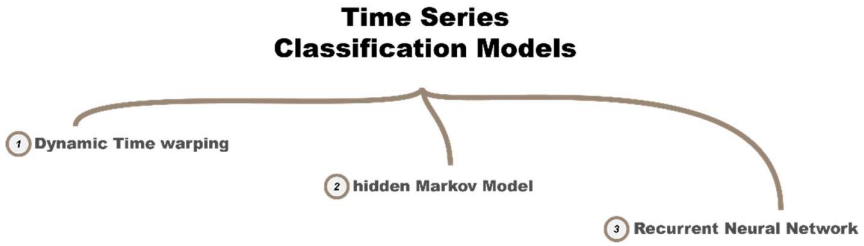


Figure 5.1 Core approaches in time series classification

5.1 Data gathering

When a feature representation is developed, such as a skeletonized image or features derived from this, data needs to be retrieved. In doing so, to be valid, the data needs to resemble as close as possible to the real live setting in which the future classification will take place. Time series data can, due to these conditions, be distorted or contain noise. Here several ways of dealing with these distortions will be discussed. These pre-processing steps take place prior to the classification of a signal. Besides pre-pressing the data, formatting and database creation should be considered. The database can be directly integrated into the platform or can also be a separate entity for sharing and labelling purposes.

5.1.1 Dealing with noise

As some characteristics about the data are known, noise reduction can be applied in several ways. The characteristics of movement in any kind of exercise inherit shape similarity to natural shapes such as Gaussian (sigmoid, and others) curves and sine waves. In the domain of displacement Gaussian (and sigmoid) like curves in time occur due to the changes within muscle power during contraction (Figure 5.2) as well as starting and ending positions of a movement are associated with close to zero speeds. This means that there will be initially acceleration and eventually deceleration. In terms of the displacement in time this means that derivatives increase and decrease gradually over time. Sine wave like will occur in reparative motion such as circumduction and in acceleration paths of other reparative motion.

As these shapes can be predicted, curve fitting can be applied to segments in a signal. Curve fitting can be algebraic or geometric. In an algebraic approach, typically least squared error in one dimension re used to optimize a fit. Where geometric fitting tries to optimise on global least squared error where multiple dimensions can be considered. The Gauss–Newton algorithm is an algorithm that can be used in performing curve fitting. With missing values or estimation of a general shape within a movement, Gaussian processes can be used as it models the probability distribution of a set of time series data at each given time. With these smoothing techniques, some data can be lost that can provide critical information about the quality within a movement. A tremor could for example increase the squared error within a fit. Therefore, the squared error on itself could be considered during classification, as in, it can be an additional feature.

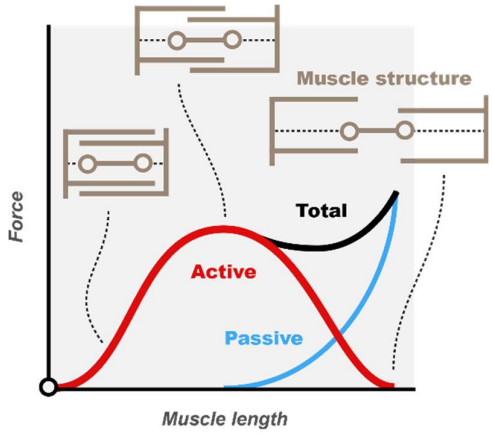


Figure 5.2 Muscle force-length diagram showing the restrictions that with permeate within the expectancy of motion signal path characteristics.

5.1.2 Model parameters

Machine learning models can be categorized on all kinds of different attributes. The most distinctive/impactful in the data gathering is the supervision in learning. Supervised learning enables models to categorize based on labels provided in addition to the data. Models like neural networks, support vector machines and trees can make use of this to create decision boundaries. Unsupervised learning tasks include clustering techniques where parameters need to be provided such as number of clusters, shape of distribution and in time series the optional the cluster appearance hierarchy. The number of clusters can be estimated using techniques such as Bayesian information criteria (BIC) that penalizes more parameters in models or least absolute shrinkage and selection operator (LASSO)/ridge. For anomaly detection, techniques such as Ranzac can be used. Multiclass classification can categorize a signal in most likely class (walking vs running e.g.) and regression could provide a scale (quality measure). Furthermore, whiles training a model parameter values can differ based on the initial conditions of the parameters. This can be referred to as an optimization problem and is tackled with training multiple models with varying initial conditions so that the solution space is extensively explored. Basic techniques include cross-validations, more sophisticated techniques let models' parameters coevolve with bio-inspired methods such as particle swarm optimization and graphitational search algorithms (error rate influences an update of parameters in other models).

5.2 Starting pose

In classifying or simply analysing sequential data that belongs to a class, data needs to be clipped. This means that fragments in continues interaction should be extracted, distinguishing idle movement from actual exercises. This can be achieved in multiple ways where a manual clipping is the most straightforward. Manually a starting/end can be determined by the user by simply pushing a button. A system fixed countdown and automatic stop comes close to this manual approach. Both these manual approaches on staring/stopping lag a flexible view. Extra idle information can, with these approaches, be incorporated within the stored segment. A fixed time can lead to incomplete recordings when a subject did not execute the exercises in an expected paste. For classification and data storage purposes more, sophisticated techniques are preferred. Such that start, and end are automatically detected. This could be simply a learned threshold value of one or multiple feature values (Figure 5.3). In addition, at the start of an exercise there should be a check if the correct initial pose within the user is present. This can simply be a sub feature space where a check can quickly decide if the exercise can be started in a correct manner.

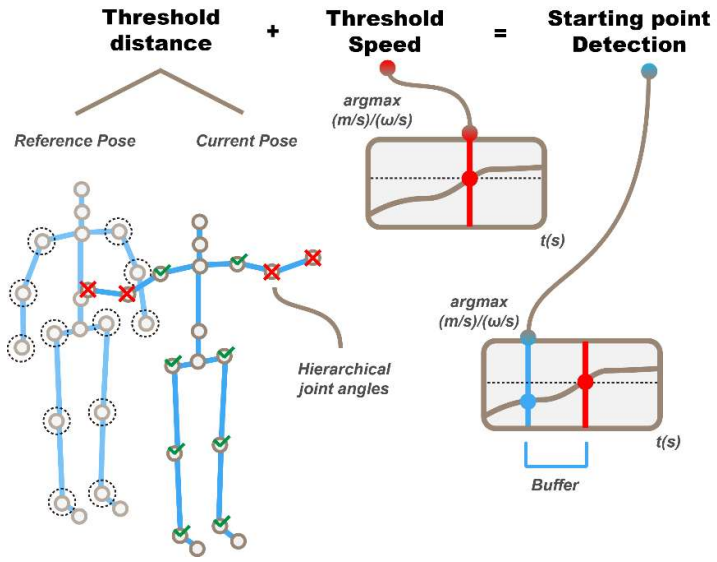


Figure 5.3 Starting pose detection, with joint specific position and movement thresholding.

5.3 Exercise Taxonomy

As is discussed in [57], recognition can be typically a single-layered approach which means that a sequence is analysed at whole (space-time trajectories, motion history images) or hierarchical that include fixed syntax and is translatable to descriptions, e.g. first step to the front and then make a step to the side. As in [11] the use of target poses can be used in classifying the sequential steps within an exercise and therefore pinpoint in a descriptive fashion where errors occurred. Hierarchical approaches can aid in guiding the user to execute partials in the correct order. This would especially be useful when exercises become more complex and longer. For shorter exercises, hierarchy can be useful to analyse the target features. As discussed, the paths (in this case every syntactic part) can be generalized to a descriptive curve with a certain length in time. This curve can be restricted so that the assumption is made that path shape does not drastically change with increasing speed or range of the movement and the squared error to this path can be added as a rate of error. Another approach is to apply curve fitting and use the curves parameters to determine a rate of error. To be able to cut up an exercise into subtasks, methods of key point extraction are needed. These key points could detect extrema and minima within range and speed for example to split an exercise into upwards movement and downwards (Figure 5.4).

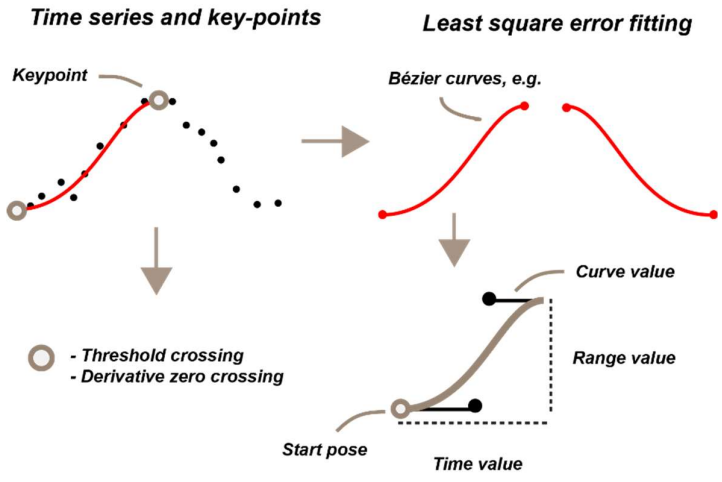


Figure 5.4 Abstracted curve fitting proposal

5.4 Dynamic Time Warping

This technique is used to align sequences that are variant in time onto each other. A reference signal needs to be choses, in this case a correctly executed exercise could be chosen as this reference. To be noted is that this reference could also be extracted from multiple examples by learning the distribution. Sequential data points will be mapped together (non-linear) so that further similarities between the sequences can be expressed. Constraining the limit and distribution of the warping can be used to detect if a sequence is sufficiently similar in time (or to detect if it is of the same class). After warping, individual differences between the template sequence and the warped sequence can be evaluated over windows and then, be classified as strongly or weekly similar. Decision boundaries that are learned from the distributions of correct and incorrect executions indicate the quality within parts of an exercise.

As therapeutic exercises are inherently stochastic not only in time but in corresponding values, distributions are better representations of the correctness. Using the distribution of a set of correct execution to create a correctness value would therefore be advisable. As in [53] these distributions are found by first aligning all executions to a sequence with a median length and then fitting multi Gaussian density functions onto the cross-sections. As DTW only expresses a similarity value, additions such as weighing features can be useful for classification. As in [58] this weighted value is defined by the contribution of a joint to the total amount of the movement.

A standard procedure while utilizing DTW is to map two segments onto each other. It is however also possible to do partial matching of a signal onto another one, ones or multiple times [59]. This is especially useful when the occurrence of a certain gesture needs to be detected. However, in the process of the tele rehabilitation as suggested, the occurrence is expected and will be clipped to its relevant parts. Therefore, only the direct mapping of two signals will be discussed here.

First a matrix is created with on each axis values of a sequence (1: N and 1:M) respectively. This creates a so-called cost matrix where at each point of the N by M matrix an absolute difference between the sequences corresponding points is calculated (Figure 5.5). Then an optimal warping path needs to be found. In this process an accumulated cost matrix is used, which adds to a position the lowest predecessor value with the restrictions that predecessors are at least of a lower value in one axis and can't be higher in any (b). Then the warping path needs to follow the valley of the least resistance. This valley is found iteratively in reversed order, as in it is found staring at the end of one sequence saving the argmin position for each sequence value, leading to a warping path.

As this technique is only a mapping tool, recognition needs to be further shaped by restrictions, maxima in local and accumulated cost values, that can be leaned with the help of distributions of correct executions (or other qualities). Clustering techniques and Bayesian inference can then provide the likelihood of an execution being correct or of alternative quality.

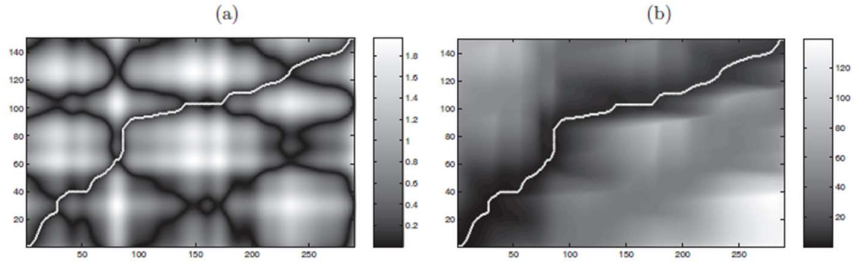


Figure 5.5 Cost matrix and accumulated cost matrix (mapping of two sequences) with in white the optimal warping path[2]

5.5 Hidden Markov Models

Hidden Markov Models (HMM) are statistical models that are well known in the domain of temporal pattern recognition. They can provide time-scale invariability when it comes to recognition of these temporal patterns. A HMM models real wold data into so called hidden states. Domains amongst witch Hmm's have been applied successfully include gesture recognition, speech and language processing, methodological forecasting, stock market/economical trend analyses. They have been used in human motion classification attempts as early as 1992 [60]. HMM is described in terms of probabilities. These are initial, transitional and emission probabilities. Initial probabilities are the distribution of probabilities of 'being in a state' before a sequence is observed. Transitional probabilities

[2] *Retrieved from [59]*

are represented by a matrix, in which the probabilities indicate the possible changes from one state to another. Finally, the emission probabilities model the variance of each state's associated values (mostly Gaussian probability density functions (PDF) obtained from continuous variable observations. These model parameters can be learned with the use of the Expectation-Maximization (EM) algorithm. Signals can be classified by looking at the probability that a signal is generated from a trained HMM. This is done using the forward algorithm. The so called backward algorithm predicts the hidden states of a given sequence of observations.

The principles of Hmm's are (I) that there is a set of real world observations $(o_1, \dots, o_n)$ where $(o_1, \dots, o_n) \epsilon X(Discrete, \mathbb{R}, e.g.)$ but these values are mealy a representation of the real world 'states'. Those 'states' can be inferred by the observations. This set of hidden 'states' $(h_1, \dots, h_n)$ will take values as following: $(h_1, \dots, h_n) \epsilon Y(\mathbb{z})$. *(II)* The joined distribution of these sets of values respect the graphical model as shown in Figure 5.6. By the factorization of its joint probability $p(o_1, \dots, o_n, h_1, \dots, h_n)$ being Equation 5.1, So that each state depends on its previous state and the observations are conditionally independent given the state.

$$p(h_1)p(o_1|h_1)\prod_{i=2}^{n} p(h_i|h_{i-1})\, p(o_i|h_i)$$

Equation 5.1 Factorization joint probability of a sequence

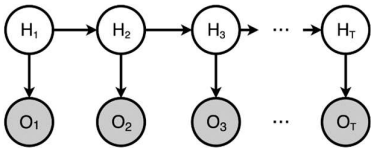


Figure 5.6 A Trellis diagram, the graphical model of an HMM [3]

Transition probabilities: This includes all the probabilities of changing from one hidden state (i) to another (j) $T_{(j,k)} = p(h_i = k|h_{i-1} = j)$ where $j, k \epsilon Y(\mathbb{z}(1, .., n))$ and results into a $n \cdot n$ sized matrix (Transition matrix). Emission probabilities: $\varepsilon_i(o) = p(o|h_i = l)$ where $l \epsilon Y(\mathbb{z}(1, .., n))$ and correspond to the states. $o \epsilon X(Discrete, \mathbb{R}, e.g.)$ and corresponds to the observed data. Each $\varepsilon_l(o)$ is a probability distribution (probability density function) of the observed data for state l. There are different distribution types depending on the values of X. A distribution type for these probabilities must be defined within the algorithm. For Discrete values, the distribution type could be a Probability mass function (PMF) or multinomial distribution, whereas for $o \epsilon X(\mathbb{R})$ commonly used distribution types are Gaussian based (regular Gaussian emissions and Gaussian mixture emissions). Initial distribution: $\pi(l) = p(h_1 = l)$ where $l \epsilon Y(\mathbb{z}(1, .., n))$ (being a PMF). This leads to Equation 5.1 being described as following:

$$\pi(h_1)\varepsilon_{h_1}(o_1)\prod_{i=2}^{n} T_{(h_{i-1},h_i)}\, \varepsilon_{h_i}(o_i)$$

Equation 5.2 Factorization joint probability in terms of transition, emission and initial probability distributions

Using the forward algorithm, we can calculate the probability that a sequence is generated by a trained HMM. Calculating the probably for a sequence on N trained Hmm's and choosing the highest probability provides us with a class assignment. HMMs are useful as they provide insight in the distributions of specific sets of sequences. The possibility to evaluate these distributions between different trained models can aid in development of a better understanding of the phenomena that is examined. In training (fitting the distributions on the data), the amount of states need to be specified (with techniques such as described in 5.1.2).

5.6 Recurrent Neural Networks

Different types of neural networks can be used to perform classification on time series data, examples are: RNNs such as Boltzmann machines (BM) or LSTMs. As in different research fields (speech recognition and maintenance/stock prediction), RNNs seem promising in the field of motion recognition [61]. The networks differ to regular NNs by having multiple inputs. They have an additional feedback-loop that flows as output from a node to become its own input again (so called memory), enabling the network to model contextual information. In Figure 5.7 this basic principle is demonstrated, where a whole sequence has hidden states (h0-ht) that act as this memory. A popular RNN variation, LSTM, overcome the problem of a vanishing gradient (decay of information over time) with the so-called gating technique. This technique overcomes the problem by the models extended functionality so that it can decide when input can be forgotten or when it needs to be remembered for future timestamps. RNNs can be used in forecasting tasks and

[3] *Retrieved from https://stathwang.github.io/images/hmm.png*

could therefore be useful within the field of tele-rehabilitation to predict harmful situations such as falling or harmful executions that could be communicated in advanced so prevent the occurrence. RNNs are used within different studies for human motion/activity recognition [62]. High accuracies are achieved using public available databases. However, similar movements such as bending or picking up and throw are more often misclassified. The work states that the short occurrence of the throw could be the reason of misclassification, as its impact over the whole signal is diminished. Reweighing temporal input on elapsed time or utilizing temporal classification windows could improve this. Compared with DTW and HMMs, RNNs needs more computational power to be trained and insight into the structure does not provide (generally, unless engineered for the specific reason) insight into the phenomena being learned (back-box approach). New ways of using NNs to avoid retraining when new data is available are so called one-shot learning techniques that use networks to serve as an augmented memory and are possibly a first step in a general intelligent system where models could be applied into any learning problem referred to as model-agnostic by Finn at al. [63].

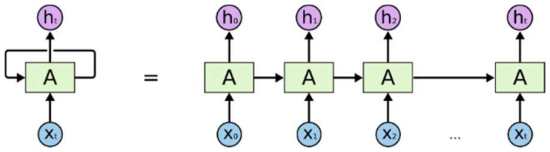


Figure 5.7 Graphical representation of a rolled-out recurrent neural network where each x represents one value that is inputted into the node at a certain timestamp. H is the output where in the rolled-out version they can be revered to as hidden states. [4]

5.7 Research perspective

The research area of human motion classification utilizing techniques as described in this chapter started as early as 1992, where HMMs where used to perform gesture recognition [60]. Besides the main approaches all kinds of different hybrids or derivatives of the classic algorithms are used to optimize performance. In Table 5.1 a compact overview is provided of these different researches during the past decades. Most classifications problems are proposed to be solved with the use of class labelled data. In only one of these researches an attempt in creating a clinical relevant score shows segmentation of the exercise into joint groups that are labelled individually.

Table 5.1 Classification methods in human motion recognition, experimental developments

Year	Classification method	Classification output	Features	Ref.
1992	HMM	6 gesture assignment	Downscaled binary image sequence	[60]
2000	Switching linear dynamic system (SLDS)	Binary classification	Video data	[64]
2008	Hierarchical Aligned Cluster Analysis (generalized DTW kernel)	Gesture segmentation	variable number of features per cluster (state)	[65]
2012	hidden Markov models	8 and 10 gesture assignment	LDA, clustered into k posture visual words (joints) view invariant	[66]
2013	Weighted Dynamic Time Warping	8 gesture assignment	Joint representation	[58]
2014	Naïve-Bayes-Nearest-Neigh-bour (NBNN)	16 gesture assignment	EigenJoints (Accumulated Motion Energy in skeleton data)	[67]
2014	probability-based DTW	8-15 gesture assignment	Bag-of-Visual -and-Depth-Words	[53]
2015	5 bidirectional recurrent neural networks (BRNNs)	65 gesture assignment	Joint groups internal orientation (5 in layer 1, 4 in layer 2, 2 in layer 3, 1 in layer 4 and 1 in layer 5)	[61]
2016	Hidden Semi-Markov Model (HSMM)	Clinical score (range 0-100), Likert questionnaire about target and posture of 7 body segments.	global movement descriptors (Kinect joints to angles)	[68]
2016	Generalized Canonical Time Warping	Similarity value (like DTW)	Video (landmarked)	[69]

[4] *Retrieved from http://colah.github.io/posts/2015-08-Understanding-LSTMs/img/RNN-unrolled.png*

6 Requirements

The requirements in assessing the quality of the therapeutic exercises that will be integrated into the rehabilitation plan can be split in user, system, feature, model, and experimental requirements (Table 6.1, Table 6.3). The latter one ensures that the assessment can be projected onto further work and provides the guidelines within this thesis on how and which experiments are going to be performed.

6.1 User requirements

As the different users have varying interest whiles users the platform, features, models and integration of the assessment tool will need to be broadly customized accordingly. Within the assessment, the influence of the specialists that perform surgery and those who do post-surgery care will have the least influence. Their only requirement of influence is that harmful situations can be detected and if possible, be prevented. The physiotherapists will have the most influence as the data and its representation needs to be understandable in their jargon and be manually assessable. This means that the architecture of the feature structure and weights within the model will be founded on the general assessment methods that are currently used. The models will need to be able to segment exercises as localization of errors within a movement is important (imbalance at the end needs to be labelled different as insufficient range of movement e.g.). Furthermore, consistency and improvement needs to be assessable with the classifications that are made to support the therapist in pinpointing the stage of progress of the rehabilitation. The patients only concern is that the feedback has enough descriptive power and will be provided at the right moment. Arguable the right moment can differ between exercises, this could be a live feedback loop when the exercise is relatively long or when timing matters and after the compete movement when those factors differ. One major variant is which hip needs to be rehabilitated (left/right). Therefore, a mirroring function should be implemented into the assessment tool. For a more general-purpose, target rehabilitation areas could be set and initial patient status to ensure that feedback is tailored user specific.

Table 6.1 Acquired requirements of caregivers, Therapist and patient viewpoint

<table>
<tr><th>Caregivers</th><th>Therapist</th><th>Patient</th></tr>
<tr><td rowspan="3">Harmful movement detection</td><td>Jargon translatable feature representation</td><td>Mirror possibility</td></tr>
<tr><td>Progress estimation (variance and distance to goal)</td><td rowspan="2">Receive feedback at the right time (real-time)</td></tr>
<tr><td>Localize error in exercise</td></tr>
<tr><td rowspan="2">Harmful movement prevention</td><td>Data gathering</td><td rowspan="2">Custom goals (cognitive/ physical)</td></tr>
<tr><td>Model creation</td></tr>
</table>

6.2 Platform Requirements

Within the platform, the assessment tool should be able to operate at a remote location with respect to the user. This means that on the platforms server, the assessment will be executed, and feedback will be send back to the user. A clear description should be provided within the assessment tool of what the input and output values should represent. This include the transformations, structured naming of exercises and models that will be used per exercise. Within the way training of the models will take place, the assessment tool should automatically name models correctly and save them in the right directory.

6.3 Features for tele-rehabilitation

The features in this process need to compromise between descriptive power and required memory storage. As the database over time will grow, the data needs to be represented in a compact form. The data will be safeguarded to retrain models, to be able to manually examine executions (additional labelling) and build up a mocap database that could be shared in scientific communities.

Skeletonized data will be used as this captures the pose in a compact form and as discussed has an advantage whiles executing classification tasks (section 4.5). For the classification, the features additionally need to be invariant to body shape and as much as possible to the position of the subject with respect to the recording device. Some transformations will inherently obey this, others will need to be normalized with respect to dimensions of certain limbs of the subject (speeds e.g.). Therefore, more general features are required that expresses movement in a similar way as therapists themselves do (Figure 3.3). Expressing motion of the different joints into the relative pose of the connected limb would therefore be suggested (angles). In all cases the movement should be expressed in terms of the perspective of the user, meaning a personal coordinate system is required (defined by 3 planes, Figure 3.2). This means that the location of the knee joint relative to the hip determines the amount of abduction (adduction) and flexion (extension) is occurring with the corresponding upper leg.

6.3.1 Approaches

Expressing movement into a representation accordingly to the earlier mentioned terminology is not enough. This will not capture any absoluteness of the movement. E.g. is the hip rotating the trunk or the leg. Therefore, speeds of the appropriate joints need to be considered too. The problem is that without a personal reference these so called absolute movements are still variant to the orientation of the user relative to the camera. A solution would be suggested in terms of creating a local coordinate system that will allow for an explanation of the planar personal direction of the joint movement. As speed correlates with rotation, the limb lengths do influence the eventual speed of connected joints. Therefore, speed as feature needs to be normalized on limb length. An overview of suggested featurizations are expressed in Table 6.2.

Table 6.2 Different features and their expected complexity expressed in levels

<table>
<tr><th>Fundament</th><th>Medium-Level</th><th>Advanced-level</th></tr>
<tr><td>Joint displacement (+speed/acceleration) *</td><td>Peak (ROM) detection</td><td rowspan="2">Coordination</td></tr>
<tr><td>Joint abduction (+speed/acceleration)</td><td>Ball-socket joint rotation (pivot)</td></tr>
<tr><td>Joint adduction (+speed/acceleration)</td><td rowspan="2">Spine rotation (pivot)</td><td rowspan="2">Force*</td></tr>
<tr><td>Joint flexion (+speed/acceleration)</td></tr>
<tr><td>Joint extension (+speed/acceleration)</td><td rowspan="2">Compensation detection*</td><td rowspan="2">Personal coordinate system</td></tr>
<tr><td>Joint angles (Hinge joints)</td></tr>
</table>

* Normalized on mean Anthropometric Dimensional Data

As Kinect's Software Developers Kit (SDK) creates possibilities to develop applications that utilizes the skeleton data (spatial location of joints), new representations emerge that provide useful information within the domain of rehabilitation. These new representations allow the detection of relative body part orientation/rotation and force plots on a virtual body of a current state (Figure 6.1). As the Kinect is used in animation the joint rotation is additionally provided for animators to easily apply the data on virtual skeletons with different body part ratios. This force plotting application uses models that are derived from the field of classical mechanics to project static and dynamic forces onto specific body areas. These new promising and useful applications are however not available as part of an extended SDK yet.

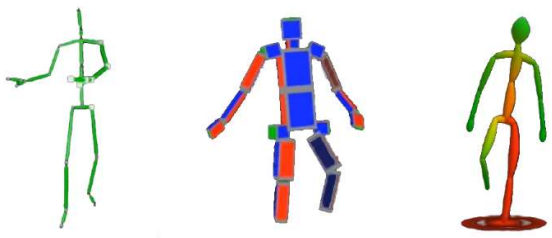


Figure 6.1 Advanced methods of depth motion representation, left joint position, middle a limb orientation and right a muscle force estimation.

6.4 Model requirements

As explained earlier, the model would preferably create the possibility of fault prevention. This means that the model needs to create real-time assessment and have the power to perform forecasting. The platform will in practice gather data from patients that can be used to create more accurate models over time and therefor, better feedback. This means that models capable of dealing with less data are preferred during the early stage of the implementation. In addition, the feedback can slowly increase as more data provide more evidence to do so. This only begs the question on how this newly data can be assigned to its specific quality. Here the therapist could re-enter the loop as a mechanical turk [70] to label ambiguous data. Collecting sufficient expert knowledge on the quality can show the consensus of the therapist on an execution.

6.5 Experimental requirements

To find a suiting method in assessing the quality, experiments need to be conducted. The first step in this process is the data gathering. In the gathering of the data, the quality should be monitored by an expert. In this sense, correct representations can be created of the feature distributions where the exercises should still be reviewed as correct (or incorrect). As the experiment can be strictly coordinated, the staring position and starting moment of the exercise can be guided so that obtained data is clear of varying initial conditions, those that might be undesirable (section 5.2) or can influence the assessment as such. Initially the subjects performing the exercises can be healthy persons of differing ages. In the firsts experiments the quantity of recordings should shape an image of the exercise ontology and variances. Eventually how this differs from real patients and a proposal on how these pre-trained models can be translatable to real patients is the objective of the total set of experiments.

Table 6.3 Acquired requirements on platform, feature, model and experimental design

Platform	Features	Models	Experiment
File formatting	Jargon translatable	Handle smaller amounts of data	Assessment method (therapist), what to asses
Data stream management	User normalized	User creatable	Used exercises
Online/offline assessment	Position to camera Invariant	Forecasting	Plausible fault schema
Saving assessment results	Projectable (avatar)	Used to simulate new exercises	Clean (data) in time and initial pose
Instruction manual	User creatable	Generalize for multiple users	Generalize to patients

Methods

7 Methodology

As initially stated, being able to measure quality is hypothesized to increase the overall rehabilitation program. The suggestion that appropriate guidance (automatically, knowing the limitations of the user) and insight in the progress (understandable for therapists, to provide useful feedback and intervene if necessary) will increase the rehabilitation, makes it that finding a suitable quality extraction method for therapeutic exercises will be the focus of the methodology. The intent to create a low-cost solution will compensate in the resolution of the expected quality measurements. Hence, with the execution of experiments, these limitations could inherently shape the quality extraction method. The experiments should provide answers on dealing with exercise data, from gathering to the assessment (keeping in mind the requirements as presented). Further, the manner of assessment by therapists in real time can be observed and will aid in a better understanding of the specialist's eye. An important aspect is the experimental replication possibility, as it expected that experiments can be spread out over multiple sessions, various locations and different therapists could aid in the process and need to grasp the procedure. Reserving a suitable location and finding test subjects to conduct the experiments is outsourced to the therapists affiliated with the project (Arián Aladro Gonzalo and Danilo Esparza). A representation of one of the recording session can be seen in Appendix D. In each experiment the methods will be explained in terms of how the classification is executed, the subject specification, exercises specification, the therapist labelling technique, the feature representation used, the hardware and software setup. The general pipeline in conducting the experiments is sketched in Figure 7.1.

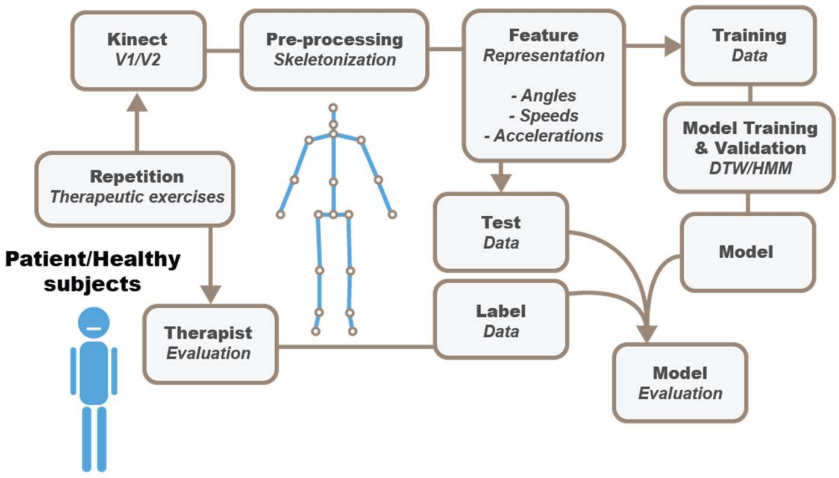


Figure 7.1 The basic methodological layout for all the experiments

7.1 Introduction

A total of 4 different experiments are executed. In every experiment new insights are created caused by the subject's understanding of the exercises and handling of feedback, therapists assessments (interpersonal) and feedback, types of exercises and their errors (severity), the exercise feature representation's descriptive power, classification interpretation (linked to therapist assessment) and environmental influences. With every experiment the setup is morphed into a more realistic version (real wold application). Within the different experiments the question that are bagged to be answered evolved as following:

Experiment I: Can we distinguish a good performed exercise from a bad one?
Experiment II: Can we distinguish several types of bad executions (different compensations)?
Experiment III: Can we detect coordination issues?
Experiment IV: Can the method be projected onto real patients?

In an earlier stage of the project the choice for a vision based system and not a wearable system is made [71] for the data gathering task. This 3D visual system (Microsoft Kinect) permits an automatic skeleton extraction of up to six persons in the image. As earlier discussed this Skeletonization shows to be advantageous in classification of gestures and would therefore be used. To record exercises, an application is created. In this application (Figure 7.2, Appendix A), the exercise name can be easily assigned, and recoding creates csv files with accumulating file number (Abduction_1, Abduction_2, e.g.). When the recording button is pushed, a sound indicates the start on which the participant could act. In addition, as a therapist might step into the scene to demonstrate an exercise or correct the participant, the true exercise subject is tracked and only their skeleton is saved into the csv file. First the raw data (x, y, z coordinates of each joint) is gathered and subsequently transformed into desired features.

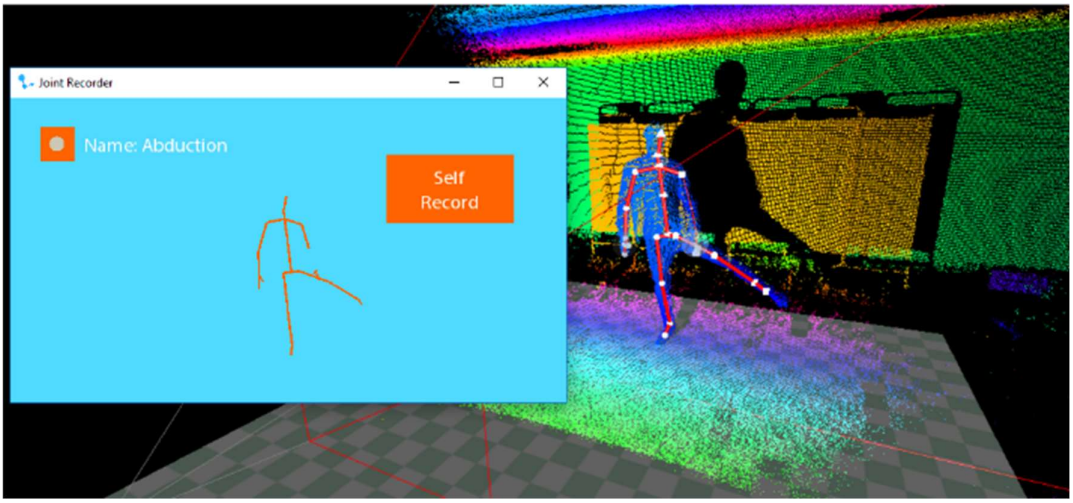


Figure 7.2 Recording application utilized in the experiments to save raw coordinate per joint into csv files

7.2 Experiment I – DTW Binary Classification

The main desire at first is to be able to distinguish compensation movements within therapeutic exercises. A first experiment utilizing DTW showed that using Kinect's skeletonized data permits high levels of distinction of a correct or incorrect execution [72]. Here an overview of this work is provided as it is the fundament of the later experiments. This work is created by the head of project Yves Rybarczyk, where the practical work is carried out as part of this thesis. This means the experimental setup, execution of the recordings and data structuring, where personally the main tasks during this experiment. An of the shelf DTW plugin is explored before conducting this experiment but did not seem to work adequately (Appendix E).

Can we distinguish a good performed exercise from a bad one?

7.2.1 Introduction

For movement assessment, the signal of the movement to be assessed is compared to the signal of the movement correctly executed (correct trials). A DTW analysis was used to assess this correctness. This is a well-established method in finding similarity between 2 temporal signals. With this similarity (cost matrix) value a discriminative algorithm can distinguish between classes, in this case correct or incorrect. The main classifiers that implement this type of method for activity recognition are: k-Nearest Neighbour [73], Support Vector Machines [74], the Naïve Bayes [75], and C4.5 Decision Tree [76]. The last two are by far the most popular algorithms, because they generally enable a high classification accuracy [77].

7.2.2 Classification

Here, the Naïve Bayes classifier is chosen because we make the assumption that all of the features (coordinates X,Y,Z of each joint) contribute equally and independently to the decision [78]. The skeleton retrieved from the image is constructed out of 20 joints which all have X, Y, Z coordinates. This raw representation (total of 60 features) is used, where recording conditions with respect to subject-camera orientation are kept equal. Thus, minimalizing the variance within the coordinates initial values for each independent trial. There are two hypotheses (H) for the movement assessment: correct vs. incorrect. To get a probabilistic value (between 0 and 1), the likelihood of each hypothesis (or class) is normalized. The main possible issue in using Naïve Bayes method is in case of redundant attributes. In this situation, it is possible to use additional methods for feature selection, to select a subset of independent attributes.

7.2.3 Protocol

Eight subjects participated in the experiment. They were asked to execute four different rehabilitation movements. Each movement was repeated eleven times: six times correctly and five times imperfectly. The correctness of the exercises was labelled by a physiotherapist. The first movement was a hip abduction (HA). The second movement was a hip extension (HE). The third movement was a slow flexion of hip and knee (SFHK). And the last movement was a sequence, in which the subject had to do one step forward, one step sideways and one step backward (FSB). These movements were performed on the right side, only. For these rehabilitation exercises the main mistakes that an individual performed were: (i) an inappropriate amplitude of the movement (too short or too large), (ii) an additional flexion of joints not involved in the exercise (e.g., trunk flexion), (iii) an execution of the movement in the wrong spatial plane, (iv) an incorrect positioning of the centre of mass. These errors were used as imperfect trials of the experiment. During the execution of the movements, subjects were in stand-up position and at approximatively 2.5 meters from a Kinect camera (Figure 7.3). The Kinect height was aligned with the xiphoid apophysis of the subjects. A program was developed to record the 3D-coordinates (X, Y, Z) of each joint of the Kinect Skeleton (Figure 7.2). Thus, twenty joints were analysed. The framerate of the motion capture was 33Hz, approximately.

Figure 7.3 Experimental environment where the subject stands in front of a wall facing the Kinect

The classification is based on a personal reference signal. This means that a performance is measured by using a correct labelled reference signal of that same subject. As each subject has 6 correctly labelled executions, firstly one is assigned to be the reference of these 6 and then the similarity is calculated with this signal for the remaining 10 (5 correct, 5 incorrect). The reference signal then is swapped with another correct execution and another 10 similarity values are created. This is done until each signal labelled as correct has been the reference signal and created similarity values for the rest of the executions. This creates a total set of similarity values that are equally divided into the classes correct and incorrect.

Not all the sixty attributes are essential to proceed with the assessment of the movement. Thus, a selection of the relevant attributes for each kind of exercises could be performed, to improve the classification (correct vs. incorrect) of the movement (elimination of redundant features) and to get a simplified model of assessment of the correctness of the movement (only based on the most pertinent features). Several techniques are available to automatically perform this selection. The method used in this study is the "wrapper" attribute selection [79]. This method can be applied backward, forward or bi-directional. The backward searching consists of removing one attribute (the worst one) at each search step. On the contrary, the forward searching start with a zero-attribute subset and add the best attribute each time. The bi-directional is a combination of backward and forward searching. In all the cases, the search stops when the classification performance gets worse. The "wrapper" method uses cross-validation to select the best attribute to add or to drop at each stage. To sum up, two components must be defined to apply this technique in practice: a search method and an attribute evaluator. The search method defines the searching direction and the search termination criteria. The attribute evaluator evaluates feature sets by using a learning scheme and classifier. In this study, the setup used is backward searching and Naïve Bayes classifier.

7.2.4 Results

Experimental Results - Overall, the average classification accuracy of the movements is 98.2% (SD = 1.1). The mean of accurate assessments for each movement is higher than 97%. Table 7.1 shows that the accuracy of this classification is almost the same between the different exercises and the different subjects. These results suggest that the DTW is an appropriate technique to discriminate between correct and incorrect execution of the rehabilitation movements, if so a personal reference signal is used.

Table 7.1 Percentage of accuracy in the assessment of the movements.

Subjects	HA	FSB	HE	SFHK	Mean
1	.96	1	1	1	.99
2	1	1	.98	.96	.985
3	.91	1	1	.93	.96
4	.96	1	.98	.98	.98
5	1	1	.96	1	.99
6	1	.96	1	1	.99
7	.98	.93	1	.96	.968
8	1	1	.96	1	.99
Mean	.976	.986	.985	.979	.982

After feature selection the overall percentage of accurate classification of the movements is 98%. As previously, none of the movements have a classification lower than 97%. The similarity between the results with and without attribute selection is confirmed by a T-test analysis that shows no significant differences between the classification performance on these two datasets (p = .7). It is to note that the value of the standard deviation is slightly lower when a selection of features is performed (SD = .71). This fact suggests that the inter-individual differences in the assessment of the movements are reduced when they are based on a selection of the most relevant attributes for each exercise and individual. In addition, it is preferable to build a model based on a selection of the most relevant attributes than to use the whole features, because we will get a simple model as accurate as a complex one. This characteristic will be fundamental when the model will have to assess the correctness of the movements in real time.

Different models were created for each subject and exercise through a feature selection based on a wrapper technique. Table 7.2 shows a synthesis of the main joints involved in the assessment of the rehabilitation movements. The main features that enables the algorithm to discriminate between a correct and incorrect movement are related with the "Right Foot", the "Right Ankle" and the "Right Knee". This is not surprising considering that all the exercises asked to the participants were designed for the rehabilitation of the "Right Hip". Taking together, these three joints represent 78% of the features used for the assessment. The other 22% are represented by different kind of joints according to the exercise and the individual. With 62.5% of "Other Joints", "Hip Abduction" is the movement with the largest inter-individual variability. On the contrary the "Forward, Sideways and Backward" sequence is the only exercise that can be exclusively assessed on the base of the lower limb (100%) involved in the movement (mostly foot kinematics). Also, "Hip Extension" and "Slow Flexion of Hip and Knee" are mainly evaluable through an analysis of the lower limb in movement with 80% and 90% of the recognition based on these joints, respectively.

Table 7.2 Percentage of the joints used to assess each exercise.

Exercises	R Foot	R Ankle	R Knee	Other Joints
HA	.375	0	0	.625
FSB	.625	.1	.25	0
HE	.4	.2	.2	.2
SFHK	.5	.3	.1	.1
Mean	.48	.15	.15	.22

7.2.5 Conclusion

This work shows that using a DTW approach is successful in discriminating correct from incorrect executions and by doing so only a minimal number of features are needed. DTW can also be useful in segmentation (this part of the exercise is an idle gesture) of recording data as a data structuring tool. The downside here is that unprocessed skeleton data is used, which makes this approach highly sensitive with respect to interpersonal differences, camera positioning and deviations in the direction of movement. In addition to the feature representation, the discriminative approach leads to the need of gathering sufficient data. This might be an issue whiles implementing the platform, as the personal data at the start of the implementation phase will be limited and not of the expected reference quality. Also, the translation of segment analysis (this part of the exercise is wrong, e.g.) could be difficult utilizing DTW. Here a translation could provide the therapist with insight on structure where parts can be represented as for example: upwards movement of the leg or extension of the knee.

7.3 Experiment II – HMM Compensation Classification

The second experiment is set up to study the applicability of HMM models in assessing quality in real-time. This quality here is explained as being of a correct or compensatory nature, with compensation categories that cover the most common compensation behaviours during regular rehabilitation. These types of compensations will be further discussed in the next section.

Can we distinguish several types of bad executions (different compensations)?

7.3.1 Introduction

The previous experiment shows that DTW is able to assess the quality of therapeutic exercises into good and bad with high accuracy [72]. Calin [80] describes a comparative analysis between DTW and HMM for gesture recognition using both Kinect V1 and V2. Although obtaining a high overall accuracy with DTW, the study points out the fact that the performance of the algorithm is very sensitive to the database size. Addition, the authors claim that it is preferable to use HMM than DTW for gesture recognition, because it enables the system to be dynamically created and adjusted (desirable within the platform, likewise one-shot learning algorithms that can be Bayesian or Neural networks). Unlike HMM, DTW cannot model the stochastic nature of the signals. As it is a deterministic method, there is no knowledge about the variance within a specific movement. A hard boundary decides if a movement belongs to a category or another. Some authors attempted to implement a probability based on a DTW approach [53], but this is not yet applied successfully in practice [81]. Another disadvantage is its limitation for a real-time implementation. If a signal is classified on the fly, a temporal segment needs to be matched to a part of the reference signal. This is possible, as shown in [59], but it involves an additional matching threshold, which makes it prone to errors in classifying stochastic signals, especially on small datasets. In an additional study (Appendix E) the usage of a plug and play DTW module is found to be unsuccessful in providing in time feedback. Due to the facts that a real-time assessment is crucial and the movement variations play an important role in the evaluation of the exercises, the HMM is chosen to classify motions. HMMs can be used in real-time without the limitations mentioned for DTW [82]. Considering that HMMs are generative models, it is possible to find out how the categories differ from each other (based on the distribution differences).

7.3.2 Classification

Learning the model parameters (states and transitions) by optimizing the likelihood is essential to make meaningful use of the HMM in classification. The distribution function defined by a Gaussian, Mixed Gaussian or multinomial density function, as well as the covariance type, need to be characterized prior to this process. An observation is merely a noisy and variable representation of a related state. A state is a clustering of observations that relate to a distribution with a specific mean in the parameter space. A likely state is retrieved by finding the cluster that the observation is member of. Also, the transition probabilities between states creates a sequence of the most likely temporal succession of states. Estimating the model parameters is done by utilizing the Baum-Welch (Expectation–Maximization, EM) algorithm which in terms makes use of the forward-backward algorithm and is commonly applied for classifying Hidden Markov Chains [60] [83]. The probabilities are calculated at any point of a sequence by inspecting previous observations, to find out how well the model describes the data, and following observations, to conclude how well the model predicts the rest of the sequence. This is an iterative process, in which the objective is to find an optimal solution (state sequence) for the HMM. This optimal sequence of states is inferred using the Viterbi algorithm. Also, the forward algorithm can be used to calculate the probability that a sequence is generated by a specific trained HMM, making it applicable for classification.

This classification is based on training an individual HMM per subclass of an exercise. For instance, one HMM could be trained on 'running' whiles another one would be trained on 'walking' (both subclasses of the human locomotion class). When calculating the forward probabilities of a sequence of observations and comparing the probabilities of all the HMMs, the sequence is classified as the category that provides the highest probability, as described in Equation 7.1 (where λi represents a determined model and O is a sequence of observations):

$$Class = \arg_{i=1}^{n} max[\Pr(O|\lambda_i) * \Pr(\lambda_i)]$$

Equation 7.1 Class assignment

HMM state assignment - The amount of states in a HMM is a free parameter. The Bayesian Information Criteria (BIC) is a technique that aids to define a determined number of states (based on the likelihood function) by considering the possibility of overfitting the data when the number of states increases. BIC penalizes HMMs that have a high number of states, as described in Equation 7.2 as follows (where n is the data size and s the amount of states):

$$\mathrm{BIC} = \ln(n)\, s - 2\ln(MLE)$$

Equation 7.2 Bayesian Information Criteria

Therefore, the optimal amount of states is retrieved by selecting the model with the lowest BIC score. Multiple HMMs trained with a different amount of states are evaluated by cross-validation on their Maximum Likelihood Estimation (MLE) and the previously mentioned penalizing term.

Gesture representation - A skeletonized 3D image from a Kinect camera provides Cartesian x, y and z coordinates of twenty joints. The gesture representation is chosen to be a skeletonized image as this has been shown to improve the model accuracy [46]. This representation depends on the position of the subject in relation to the camera and the roll, yaw and pitch angles of the device. The causal relationships between different joints are not captured by this representation. This means that physical constrains, such as a movement of the ankle that could be influenced by bending the knee, are not accounted for. To overcome these limitations, the joints are used to create a new representation that contains angles of multiple joints in respect to the frontal and sagittal planes, as well as multiple angles between relevant limbs. Figure 7.4 shows a graphical representation of the features in relationship to the skeleton image. Table 7.3 describes the feature vector of the joint movements according to the anatomical terminology.

In this study, the motion is defined from the following joints: ankles, knees, hips, and spine. The angles of the knees are obtained by calculating the angle between ankle, knee and hip. The orientation of the knees induced by hip activity are expressed in four angular representations, following the two opposite directions for both sagittal and frontal planes. The same method is applied to describe the orientation of the torso, by finding the displacement of the centre between the two shoulders in relation to the hips. This leads to a description of the movement into fourteen features. It is the principal representation followed by a first order and second order derivatives of these features that provide speed and acceleration of the movement. Overall, a total amount of 42 features is used. Figure 7.5 represents a diagram of the HMM's implemented in this study, in which State(t) and Observation(t) are state id and associated feature values at t time, respectively.

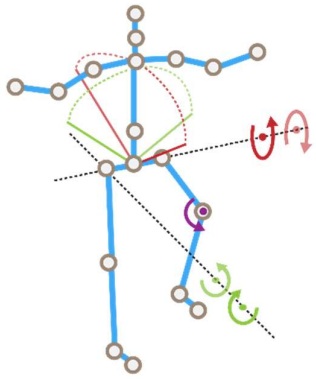


Figure 7.4 Graphical representation of the features used in the study. The movements in the egocentric frontal plane and sagittal plane are represented in green and red, respectively. The purple arrow represents the angle of the knee (independent from any plane).

Table 7.3 Feature vector describing the joint movements.
For the assessment these features are also transformed into speed and acceleration.

Right hip Frontal plane rotation (abduction)	Right hip Frontal plane Rotation (adduction)	Spine centre Frontal plane rotation (lateral left)
Left hip Frontal plane rotation (abduction)	Left hip Frontal plane rotation (adduction)	Spine centre Frontal plane rotation (lateral right)
Right hip Sagittal plane rotation (flexion)	Right hip Sagittal plane rotation (extension)	Spine centre Sagittal plane rotation (flexion)
Left hip Sagittal plane rotation (flexion)	Left hip Sagittal plane rotation (extension)	Spine centre Sagittal plane rotation (extension)
Right knee (flexion)		Left knee (flexion)

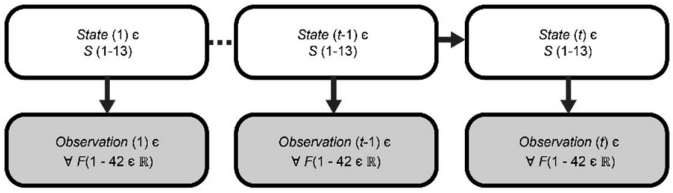


Figure 7.5 Graphical representation of the HMM for the exercise assessment. ϵ stands for: takes value out of; and ∀ stands for: out of all. F and S are the collection of features (42) and states (13), respectively. The definition of the optimal number of states is explained in section (4.1). Each state is dependent on its previous state and observations are samples of the associated current state.

7.3.3 Protocol

Four subjects participated in the experiment. They were asked to take place at approximately two meters distance from a Kinect camera. The motion capture device was placed at the height of the subject's xiphoid apophysis. Each participant executed 70 movements leading to a total of 280 records. The rehabilitation exercise was a sequence, in which the subjects had to do one step forward, one step sideways and one step backward, with variations. These variations are staged executions of errors or compensatory movements that can occur during the rehabilitation in practice. The exercise was performed in batches of ten in the following order: (I) correct execution, (II) steps too short, (III) execution without moving the centre of mass, (IV) steps too large, (V) steps with bended knee, (VI) steps with bended knee and flexed torso. The last ten trials (VII) are partially wrong executions of the exercise, in which the faults II to VI are only occurring in the beginning, middle or end of the sequence. These last executions are used to evaluate the real-time applicability of the HMM technique.

Materials - An application is created to capture the skeletonized image of the subjects performing an exercise. Python 2.7 is used to create a graphical user interface with the option to name, start and stop a recording. In addition, Python is used for the later processing steps, which are feature transformations and classifications. The application communicates with the Kinect SDK and whiles in recording mode writes the data into a CSV file with a frequency of 60Hz. The HMMLearn package for python 2.7 is used for the training and application of the HMMs [5].

Evaluation methods - Using the BIC score to select the appropriate amount of states is done for each type of trained HMMs (I-VI). It provides insight on the semantic variation within the exercises. For instance, less states assigned to a faulty movement relative to the good execution implies that 'there is something missing in the execution', whereas the detection of extra states implies that 'there is something added to the movement.

For each type of execution (I-VI) an HMM is trained, leading to a total of six distinct HMMs. To build a general model that can assess the movements of any subject, models that classify the executions of a subject are exclusively trained on the recordings of the other remaining subjects. This leads to a total of 24 trained models (6 per subject). The initial model parameters are set with a single Gaussian density function and a full covariance matrix type (initializations of model parameters are done randomly, HmmLearn standard initialization is used, EM 1000 iterations unless log-likelihood is gain is less than 0.01). The HMM topological structure is fully connected, because priory knowledge about the expected state sequence cannot be estimated with sufficient certainty. In addition, as the outcome of six classifiers determines the most likely model that is associated with the sequence, unpredicted variance in a signal (or noise) should not drastically influence the likelihood of the signal. The outcome of the six classifiers is calculated by means of the forward probabilities. Then, these probabilities are ranked from 1 to 6, where 1 and 6 are assigned to the highest and the lowest probability, respectively. A confusion matrix is used to map these values in terms of average prediction rank of each type of execution. In addition, an indication of the similarity of a type of execution in relation to the combination of all the other executions is provided. Finally, a range of sliding temporal windows is used to evaluate the real-time suitability of the approach. It is applied to assess the correct detection of the present types of faults in executions VII. These windows classify a subsequence in a fixed number of samples, which partially overlap over time. With this measure localizing the errors to specific frames can provide more insight in the severity of the error occurrence.

Validation - A 10 times repeated random sub-sampling (initializations of model parameters are done randomly, HmmLearn standard initialization is used, EM 1000 iterations unless log-likelihood is gain is less than 0.01), Monte Carlo cross-validation (MCCV) is used to evaluate the performance of each model (6 per subject and 24 in total). Results with Gaussian mixtures on real and simulated data

[5] *(https://github.com/hmmlearn/hmmlearn)*

suggest that MCCV provides genuine insight into cluster structure [84]. This is a method to select the most appropriate model for classification. To assure the ability of the model to generalize well, the validation is executed by applying, for each subject, the other three subject's recordings. This means that each trained HMM is used as classifier of the data of an unrelated subject. Each fold contains a trial of the three subjects. The split (80% train, 20% validation) is newly created during every validation (10 times) with a random assignment of the trials in the training and test sets. This leads to a model trained with 24 exercises (8 of each subject). To perform the random assignment, the python built-in random function that implements the Mersenne Twister regenerator method is used. During each validation, 6 HMMs (models I-VI) are trained such as the best performing (based on the validation score) set of HMMs is selected as models set for classification. Forward probabilities are calculated for each HMM. When the correct HMM outputs the highest probability the classification value becomes 1 and contrary 0. Per fold, each HMM classifies the remaining 6 exercises where the performance per fold is the fraction correctly classified exercises (sum of classification results) of the total classifications (36) of the 6 HMMs combined. The model's parameters differ slightly between the sets as the random data selection alters the learned state Probability Density Functions (PDFs) per fold. The best performing model set out of the 10 validations is then selected to perform the classification for the test subject.

Optimization methods - Optimization takes place in case the classification does not result in high classification performs. It needs to overcome the model's incapability by uncovering HMM specific states/features space that are associated to non-ambiguous characterization of the HMM. Each HMM learns a correct representation of a movement, but does not provide class distinguishing information. This information can be revealed by means of inspection of the distribution overlap. First, the predicted state occurrence in the classified sequences is characterized by the Viterbi algorithm (this algorithm predicts a state sequence given an observation sequence). This leads to a percentage value of each state occurrence per HMM (Pseudo code 1, Appendix B). Second, the Monte Carlo method is used to approximate the overall distribution by means of generated data draws from all HMM states. From each state, its percentage times 10000 from the PDFs are sampled.

Then, each dimension (feature) can be inspected on sample overlap in a histogrammed fashion (30 bins with range min/max sampled values of evaluated HMMs). The overlap value per feature is calculated between two different HMM samplings, expressed as a ratio (Pseudo code 2, Appendix B, i.e. number of samples of one HMM compared to samples of another one where the denominator is always the greatest value). The ratio only counts for those bins that contain at least 1% of the sampled data, since probabilistic outliers do not always occur in an exercise and, therefore, could not be a class separator. Finally, the features with the lowest average ratio are used to determine the sample area of interest for class separation. This area of interest is set to be the area of the 50% most distinguishing bins. Samples falling outside of this sampling area are not considered when calculating the forward probabilities. Applying more feature value restrictions leads to less data usage in the classification.

7.3.4 Results

State assignment - The MLE for each HMM up to twenty states is used to define the BIC scores against the amount of states (Figure 7.6). The profile of the BIC score against the amount of states is similar between the HMMs. Thus, it is possible to identify a consensual optimal amount of states at thirteen. This makes intuitively sense as the exercise is constructed out of three distinctive parts (a multiple of three is expected), plus an initial/ending part (inactive state). Thus, each part in the exercise is described by four states.

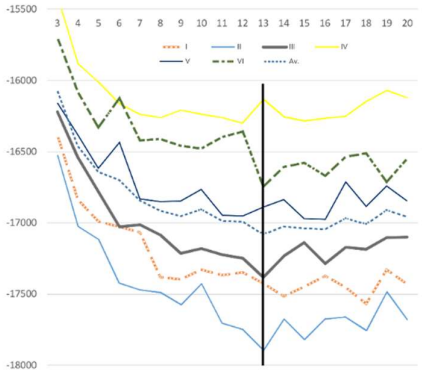


Figure 7.6 BIC scores for each type of execution (I-VI) and an averaged BIC score over these executions (blue bold broken line). The black vertical line indicates the optimal amount of states.

Classification performance - The classification performance shows a high level of accuracy (Table 7.4, left) in classifying a whole sequence into the classes (I-VI). A value of 1 means the model always gave the highest probability, with respect to the other models

and for any sequence of the related movement, whereas a value of 6 indicates the lowest probability. The values in this table are averaged prediction ranks for each model of each movement (I-VI). The average prediction rank of HMM I is the highest (2.78), which means that the execution type I (correct movement) is most closely related to all the other types. The overall performance of the classification for each class (I-VI) is shown in Table 7.4 (right). It is to note that the execution type III (i) is more likely to be classified as type I, and (ii) has the lowest prediction accuracy compared with the other classes. This could be caused by the difficulty in staging this type of execution or a lack of descriptive power in the gesture representation.

Table 7.4 Left: Confusion matrix of executions (I-VI). each column represents the types of movement and each row the output prediction ranks of the HMMs (I-VI). The closer is the value to 1 (green cells) the better is the prediction, Right: Performance of the classification of movements I-VI.

	I	II	III	IV	V	VI	
I	1	2.27	1.4	3.97	4	4	2.78
II	2.7	1	2.8	6	6	5	3.92
III	2.3	2.74	1.57	5	5	6	3.77
IV	4.74	4	4.9	1.04	2	3	3.28
V	4.27	5	4.34	1.97	1	2	3.1
VI	6	6	6	3.04	3	1	4.18

I	100%
II	100%
III	57%
VI	97%
V	100%
VI	100%

Performance optimization - To increase the overall classification accuracy of the movements, an additional data processing is proposed. The main misclassifications in the previous approach seem caused by a lack of descriptive power or a large overlap within movement I and III. Therefore, an analysis is performed to find the most distinctive parts of these movements by examining the overlap and difference of the best-defined distributions of the two most discriminative feature spaces of the HMMs, which are trained specifically on movements I and III. In this case, 'best-defined' means a feature space where HMM I and HMM III have the least overlapping samples (see optimization method). In the recorded movements, the feature combination of the angular acceleration of the torso in the sagittal plane (feature 1) and the angular acceleration of the right upper leg in the sagittal plane (feature 2) define the best feature dimensions in non-overlapping samples.

Figure 7.7 represent the 7 most prominent states (highest percentage of occurrence in classification) and the transitions between them in the feature space for HMM I and HMM III, respectively. The rest of the states are discarded in this representation as: (i) high deviation states are too general and mostly describing states that provide the function of a last resource in state assignment; and (ii) low deviation states on the other hand are too specific, which indicates a situation of overfitting.

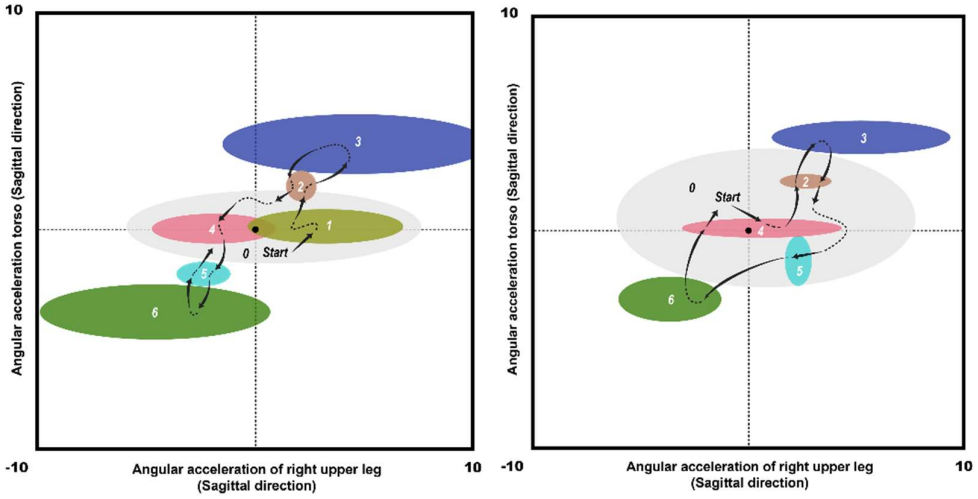


Figure 7.7 Left: Best defined states (HMM I) in the feature space described by features 1 and 2 in rad·s-2. The state distributions are visualized in terms of their first order standard deviation. The black arrows represent the most likely route of state transitions.
Right: Best-defined states (HMM III).

As shown in Figure 7.7, a similar state transition occurs in an oscillating manner, from neutral (0) to high and low acceleration, visiting intermediate accelerations during this interval. Although the same feature space is described for the two types of movement, one

state is missing in movement III (state 1). In addition, the variance of the most extreme states (3 and 6) is bigger for HMM I than HMM III and the deceleration values seem to be higher for HMM I. Thus, the main states that can clearly differentiate movement I from movement III are states 3 and 6. The other states are highly overlapping, which means that they are not contributing to the model discriminative power. There are several approaches to improve the classification at this point. The post variance of observation assigned as state 3 and 6 can be analysed and count as a weighted additional value. However, a value filter approach is used for values that repeat frequently and have a very low descriptive power (same predicted sample coverage for HHM I and HMM III). This approach is chosen as a trade-off between the critical amount of necessary observations in classification and a selection of the most discriminative values. An average loss of data that still allows an appropriate sample rate for a real-time classification (20Hz) is estimated as 60%, which is reached when applying the filter in 2 dimensions. This percentage is the basis for the filter boundaries. In Figure 7.8 the filtered region is shown as red rectangles, where the inner area is the filter area retrieved by analysing the 50% most overlapping bins per dimension. This filter excludes any observation for the classification, in which the values of the two features are <1.5 rad.s-2 and >-1.5 rad.s-2 for feature 1 (y-axis) and <3 rad.s-2 and >-3 rad.s-2 for feature 2 (x-axis). In addition, values >5 rad·s-2 and <-5 rad·s-2 for feature 1 (y-axis) and <10 rad·s-2 and >-10 rad·s-2 for feature 2 (x-axis) are eliminated, as well.

Thus, by applying the filter, roughly 60% of the data in each sequence is discarded before the reclassification. Most of these values are zeros, which causes no changes between the consecutive frames. These values could result from: (i) the variance in recording frequency, produced by the memory caches that may not cope with the short recording span per frame; or (ii) an actual undetectable movement (i.e., still body) between consecutive frames, which are all useless for classifying movement. The results show that this technique improves the classification of movement III (from 57% to 83%) and does not alter the classification of the other movements (Table 7.5). Nevertheless, the classification accuracy of movement III is still slightly lower than the rest of the movements. The main difference between movement III and the other movements is the fact that it does not involve a translational motion of the torso. It suggests that the linear movements, and not only the angular rotations of the joints, must be considered as useful discriminative features.

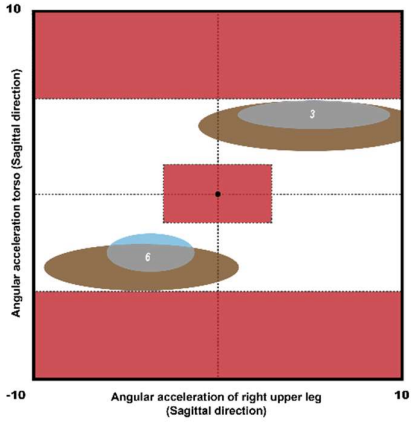


Figure 7.8 Representation of the remaining states after observation filtering. The brown distributions belong to HMM I and the grey/blue distributions to HMM III. The red squares represent areas where observations are not considered for the classification.

Table 7.5 Performance of the classification of movements I-VI after applying the feature value filter.

I	II	III	IV	V	VI
100%	*100%*	*83%*	*97%*	*100%*	*100%*

Real-time testing - In the previous section, entire executions are categorized according to the most typical compensatory movement. However, it does not provide insight regarding the severity of the execution error. An example of the lack of this insight is a low likelihood score caused by a long persisting small error vs. a short persisting large error. In other words, this section focuses on the duration, location and degree (e.g. bending knee a little or a lot) of the error. Since the objective of the platform is not only to provide an overall classification (see previous section) of the movement (correct vs. types of fault), but also to give a qualitative and quantitative assessment of the movement, a real-time classification is addressed. This classification aims to create awareness when the patient receives a feedback on phases of the movement, in which certain errors tend to occur. The result of this instantaneous classification will be displayed as a real-time feedback when the patient executes the exercise.

The samples of execution type VII, those that contain local errors within a correct execution, are used to evaluate the ability to apply the developed models in a real-time fashion. These models (HMM I-VI) provide the forward probabilities for an indefinite sequence, which would enable us to perform an assessment on a partial completion of the movement. Performing successive classifications during the execution of the exercise can disclose a switch in the class likelihood over time and, thus, localize the errors. The classification takes place over a selection of frames within the movement. Three different sizes of windows (length of the partial analyses) are used for the classification: 100, 60 and 20 frames. These different samplings are made to study the effect of the window size on the consistency and accuracy of the assessment. After classifying the frames of a determined window size, the window shifts half the number of frames in the total sequence and the classification is repeated until the end of the sequence is reached. This so called overlapping window is used to obtain a smoother classification path over time. There are multiple classification values during the full exercise. At each newly created classification moment in the exercise the values of the six classifiers are normalized in a fashion that the highest value becomes 1 and the other values are expressed as a fraction of this value. Detection is considered accurate if the majority of the movement's phase where the error occurred assigns the value of 1 to the expected error type. In the case of the execution type VII, there are three phases: step forward, step sideways, and step backward.

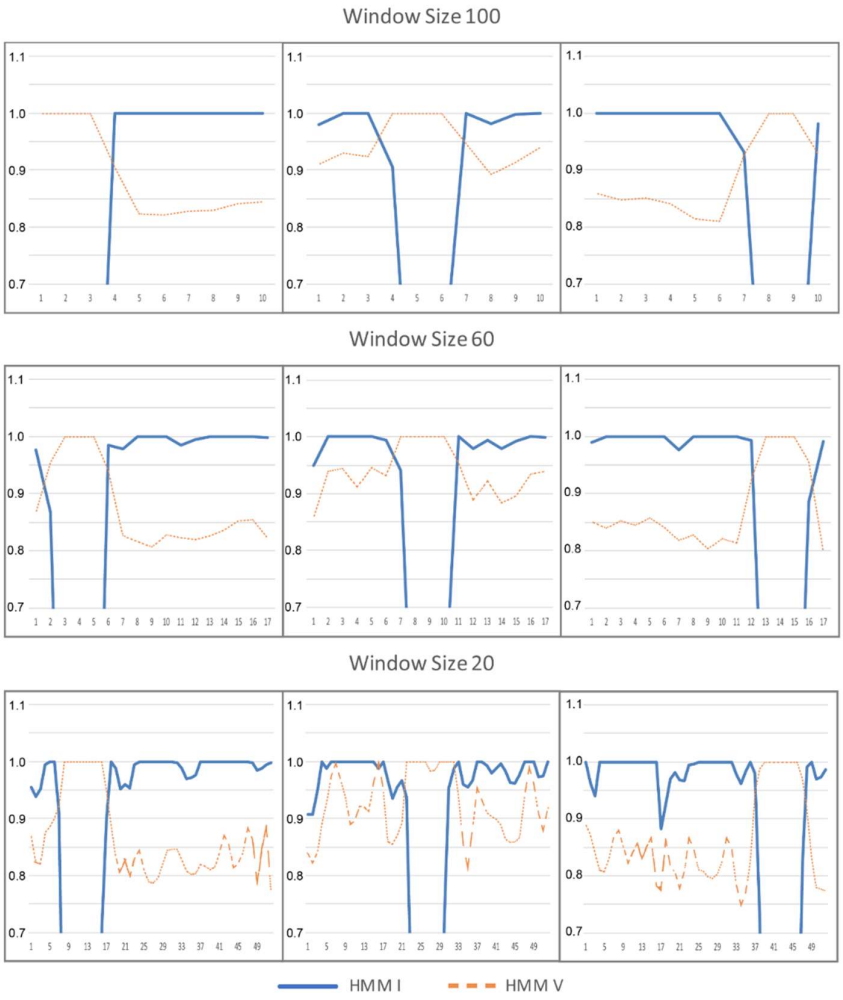


Figure 7.9 Classifications of execution of type VII where the beginning (left column), middle part (i.e., step to the side, middle column) and last part (right column) of the exercise is performed as type V. Three different window sizes are represented: 100, 60 and 20 samples. The orange dotted line represents the prediction of HMM V and the blue line indicates the prediction of HMM I (correct movement).

There is a certain trade-off for choosing the window size. A smaller window can provide a frequent feedback, but a slightly noisier prediction. Nevertheless, there is a very high detection rate (21/24) when errors of types IV to VI are present in the sequence of the movements, for any sampling size. Detecting execution types II and III are less successful (9/16). It can be explained by the fact that these two types share high similarities with execution I (see Table 7.4, left). Figure 7.9 presents three examples of a correct sequence, except in the initial, middle or ending parts, which are performed as execution V (step with bended knee), respectively. In this figure, the amount of feedback moments is represented on the x-axis and a normalized classification value on the y-axis. The sampling rate

is 50 Hz (20 ms per sample). Window sizes of 100, 60 and 20 represent approximately every second, twice a second and five times a second feedback, respectively. This example shows that the accuracy of the prediction (identification of correct vs. incorrect executions) is not significantly altered by the window sizes, which confirms the pertinence of an HMM approach for real-time applications.

7.3.5 Conclusion

This study presents a HMM approach for real-time assessment of a physiotherapeutic exercise, which will be included in a project of tele-rehabilitation platform for patients after hip replacement surgery. To be able to detect variance within movement, caused by errors or compensatory movements that may occur during the completion of the therapeutic exercise, HMMs are trained on these errors and compensatory actions. Although the setting of the experiment was controlled, the classification included intrapersonal and interpersonal variances as a model that classified a determined subject was merely trained with the data of the other participants. It suggests that the proposed assessment algorithm has a fair capability of generalization.

A high classification accuracy of the movements (97%) is obtained by building a general model that can be applied to any subject. A real-time analysis enables us to detect four out of five faulty movements, when these errors briefly occur in the beginning, middle or end of a correct execution of the exercise. The same level of accuracy is maintained whatever the detection rate (windows size down to 200 ms). These findings demonstrate that the HMM is an appropriate method to provide real-time feedback regarding the correctness of the rehabilitation movement performed by a patient. This approach is successfully applied on a real-time assessment of components of the movement, which are discriminated in several classes that differ on extremely subtle aspects. A previous work [72] has shown comparable accuracy utilizing a DTW approach. Nevertheless, this study addressed a problem of lower complexity, since it was limited to a movement classification between good and bad assessment that could be applied after the complete execution of the movement, only. In addition, the used feature representation did not account for intrapersonal differences, since the classification models were dependant of the location of the user with respect to the camera. On the contrary, the high classification accuracy and successful generalization obtained in the present study strengthens the further development of a HMM approach to assess the rehabilitation movements. The possibility to perform a real-time evaluation is a significant advantage of the HMM method, as it can provide the user with instantaneous feedback on the quality of the performed exercise. Another advantage of utilizing HMMs is that it enables the systems to be dynamically created and adjusted [80]. Since HMM is a probabilistic approach, the accuracy of the classification will increase with the individual use of the platform and the systematic update of the models.

Further work needs to include an optimization of the class separability. Robust and successful biologically inspired optimization methods such as Particle Swarm (PSO) and Gravitational Search Algorithms (GSA) have shown to create stable systems [85]. These methods can improve the performance by shaping a weighted vector of state impact in an evolutionary and robust way. This adds (i) quick insight whether multiple compensations (types of executions) can be considered simultaneously and (ii) mark possible candidate windows where errors occur (for the purpose of manual labelling of future data in the platform). In such an approach, the initial exploratory search considers the accuracy and the subsequent sequential search maximizes the classification difference between correct and incorrect HMMs. Baruah and Plamen [86] propose a dynamic evolving clustering method, in which the weight per data point evolves (decreases), losing significance as time progresses. This notion could be integrated for optimizing the real-time application, where the sliding window can be updated in a similar fashion, creating a more suitable dynamic classification.

In addition, integrating Genetic Algorithms (GA) while estimating the model parameters can increase diagnostic results of the HMMs [87]. Xue et al. [88] show that the biologically inspired optimizations PSO-HMMs outperforms both GA-HMMs and conventional HMMs. While Baum-Welch (EM algorithm) tends to get stuck in local optimum, the biologically inspired optimization methods – such as PSO and GSA – can aid in a more robust parameter estimation.

Furthermore, the descriptive power of the movements can be extended by (i) including additional features (e.g., ankle/torso displacement and normalized speed/acceleration paths of these different joints, percentages of the maximum amplitude of the movement), and (ii) creating a preliminary detection method for the recognition of noise. This noise could be caused by computational overload. Therefore, vector quantization could be applied as it can reduce the computational costs [89]. In addition, exercises can be expressed into their state sequences to learn distributions of state duration as variable parameter, which provides a further insight on the ontological structure of an exercise. State sequences are modelled in terms of duration distributions and can be used as a transition model, like in the Hidden Semi-Markov Models (HSMMs) [90]. This approach allows for a higher flexibility of the transition probability than in the HMMs, which at the end should increases the classification accuracy [91]. Finally, applying new cognitive algorithms such as Linear Discriminant Analysis (LDA) and Deep Convolutional Neural Networks (DCNN) may help to find the optimal descriptor combination to distinguish between the different classes in a non-handcrafted manner [92], which diminishes the human error in selecting appropriate features.

7.4 Experiment III – HMM Coordination Assessment

In this experiment, expanding the analytical spectrum of the HMM assessment method is the focus. As discussed in section 3.3, additional qualities within therapeutic exercises can be evaluated. In addition to the ROM and detection of a compensation the coordination and semantic structure (symmetry) can refine the feedback and monitoring possibilities. The way that these evaluations are created in this experiment follow from discussions with the therapists. They provided the insights on how these evaluations are executed in terms of their own understanding on the thought process.

Can we detect coordination issues?

7.4.1 Introduction

An automatic assessment of the quality of the performed exercises is implemented according to a HMM approach. Here, the ontology of various joints in the exercise hip abduction are evaluated from an experiment on healthy participants. This approach differs to the previous experiment as distributions per joint are learned (a HMM per joint, joint pair: Table 7.6, Table 7.7). This enables the recognition of the persistence of bad coordination. Bad coordination can be expressed as a shift of 1 joint's motion path to another in comparison to a perfect execution. In contrast to DTW the coordination value is manifested in terms of allowable transition shifts (this doesn't have to mean the sequences are similar). To be able to quickly experiment with different feature combinations and amount of HMMs that are trained for the quality assessment, an application is created where HMM models can be rapidly trained (Figure 7.10). This application is created with Python with the modules Pygame for the UI and HmmLearn for the model training. Data can be selected by pushing the 'Hmm I' button. After selecting the data, the features within the data that need to be trained on can be entered in the grey feature box. An analyses (BIC) can be performed per HMM and will be visually displayed as a graph. Then the amount of states need to be selected (manually retrieve from the BIC plot) and the models can be trained and saved (extended explanation in Appendix A).

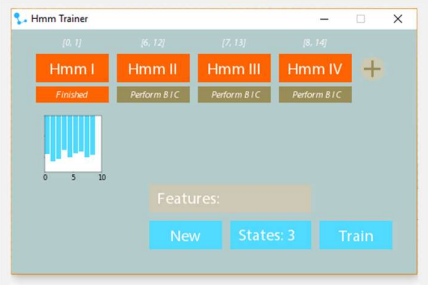


Figure 7.10 Application to train HMMs and quickly analyse BIC scores to determine states required during training

7.4.2 Classification

Gesture representation - The representation of the Kinect's skeletal data is based on the degrees of freedom per joints. The created feature vector contains 12 variables (Figure 7.11). 2 of these features are the absolute value of the linear speeds of the hip centre joint and shoulder centre joint (dark green); 4 are relative speed of the hip and shoulder that resembles axial rotation (light green). Hence, the direction of the movement of the trunk is not captured in this representation. The speeds are calculated using a buffer of 15 frames (around 0.25 sec) as a denoising method. The speed is the average (linear impact) of the values contained in the buffer. This buffer gets updated every frame by discarding the oldest data point and adding the newly one. In this sense, the data stream becomes available after the buffer being filled. The used features in analysing classification results of multiple HMMs on the movement hip abduction are presented in Table 7.6.

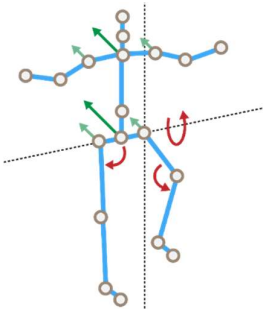


Figure 7.11 Features projected onto the used skeleton representation

Table 7.6 Feature vector used for training HMMs on hip abduction

F0	F1
Hip centre (m/s)	Shoulder centre (m/s)
F2	F3
Right Hip frontal (θ /s)	Right Hip frontal (angle)

Table 7.7 HMMs feature set and amount of states

	HMM I	HMM II
Features	0,1	2,3
Amount of states	3	4
Name	Trunk movement	Right hip frontal

HMM is a probabilistic approach that aims to model a given signal into hidden states. These states represent an arbitrary decomposition of the whole movement into successive phases. For instance, the states for hip abduction of the right upper leg are (HMM II, Table 7.7): beginning pose (state 0), moving up (state 1), hold leg up (state 2), and leg down (state 3). Here two Hmm's are independently trained on the movement of the trunk (HMM I, Table 7.7) and on the movement of the hip in the frontal plane (HMM II, Table 7.7). In Table 7.7 the feature set and the number of states used for training each HMM are shown. The states are used in the next stage to create values to assess synchronicity and symmetry. For a determined individual, HMMs were trained by using the correct data (according to the therapists labelling) of the other participants. Each model can predict the state of a specific joint or joint group. This is useful as a fault can occur in one of these groups only, whiles the rest can be correctly executed. In addition, every part of a movement can be correctly executed, but badly synchronised. Such structure enables us to identify if a movement in one joint initialized and stopped earlier or later as state transitions are easy to obtain. The characteristic pattern of a movement expressed in states leads to the knowledge of which state is associated with the movement initialization and termination.

7.4.3 Protocol

Nine subjects took part in the study. The experiment consisted of executing a rehabilitation exercise of hip abduction. This exercise was repeated eight times per participant leading to a total record of 72 exercises. Simultaneously, five therapists rated the movement on four different aspects. These aspects were range of motion (ROM), coordination, compensation and force (ratings of either excellent, good or bad). Coordination is related with the synchronicity of different body parts during the execution of an exercise. Compensations were defined as undesired movements, which were not expected to be performed. Subjects received a feedback regarding the correctness of their performance after the fifth repetition. Then, each of the three-remaining repetition was followed by a systematic feedback from the therapists. Figure 7.12 shows the experimental environment from the subject and therapist perspective.

Figure 7.12 Experimental environment where left the therapists took place to examine the subjects within a similar distance as the Kinect and right the restricted area in which the subject could manoeuvre

7.4.4 Results

The result of the performance of three subjects performing a hip abduction movement are presented here. The labelling performed by the therapists shows a global consensus, but also some inconsistences. The first subject outperformed the other 2 subjects. For subject 2 executions 2 is multiple times noted as not excellent and for subject 3 exercises 2,3 and 5 are not unanimously excellently rated. Table 7.8 shows the labels for a sample of three different participants (All labelled data of all the 9 subjects can be found in Appendix C). For each colorized horizontal strip, the first row is the ROM rating, the second row is the coordination rating, the third row is the compensation level, and the fourth row is the force rating.

Table 7.8 Labels for a sample of three subjects (one different colour for each subject). The digits correspond to the id number of the trial.

	Therapist 1			***Therapist 2***			***Therapist 3***			***Therapist 4***			***Therapist 5***		
	excellent	*good*	*bad*	*excellent*	*good*	*bad*	*excellent*	*good*	*bad*	*excellent*	*good*	*bad*	*excellent*	*good*	*bad*
ROM	12345678			12345678			12345678			12345678			12345678		
Coordination	12345678			12345678			12345678			12345678			12345678		
Compensation	12345678			12345678			12345678			12345678			12345678		
Force	12345678			12345678			12345678			12345678			12345678		
ROM	12345678			12345678			12345678			12345678			12345678		
Coordination	1234568	7		158	23467		1234578	6		3468	1257		12345678		
Compensation	134578	26		12345678	4		234568	17		34578	126		23468	157	
Force	12345678			12345678			12345678			12345678			12345678		
ROM	12345678			12345678			12345678			2345678	1		12345678		
Coordination	12345678			134578	26		12345678			2468	1357		1245678	3	
Compensation	1345678	2		1345678	2		2678	1345		2468	1357		23478	156	
Force	12345678			12345678			12345678			12345678			12345678		

The compensation (symmetry) is estimated through the symmetry of the movement, because it provides us with an insight on the motor control over the execution, which must be characterized by an evenly distribution of the agonist and antagonist muscular load in case of symmetry. It is first calculated by extracting the signal of the shoulder centre joint on both terminations: beginning and end of the rest state (0). As shown in Figure 7.13, this processing provides a middle area (shown in green, HMM I) where the minimum value corresponds to the centre of the total movement of the shoulder centre. Then, the symmetry is expressed in differences in length (or number of frames) between the left and right side in relation to this minimum value. A ratio is obtained by dividing the time took to reach the maximum amplitude where speed ≈ 0 (inversion of the direction of the movement represented by the green valley in Figure 7.13) and the time took to return to the initial pose (end of the movement). Columns 5 to 7 (symmetry) of Table 7.9 shows these ratios. Values lower than 1 means that returning to the initial pose took longer than reaching the maximum amplitude of the movement and inversely for ratios <1.

The coordination (synchronicity) is calculated by using the target movement (HMM II) and the relative shift in time of the movement of the trunk (HMM I). The shift in time is calculated by taking the midpoint of the sequenced state 2 (where maximum amplitude is reached) of HMM II and the relative difference in lengths of the shoulders clipping points on the left and the right side of this point. Figure 7.13 shows an example of an incorrect vs. correct coordination of the movements, respectively >> suggested by the therapists…. Figure 7.13 A (Example of bad coordination between shoulders and hips) corresponds to the trial 7 of subject 2, which was classified by three therapists as not perfectly coordinated (see Table 7.8, subject 2). It can be noted on this trial a minor shift of the speed paths of the shoulder centre to the left with respect to the angular speed of the hip abduction (ratio = 0.72, which is lower than the average value of 0.8 for this individual). In contrast, Figure 7.13 B shows a trial that is almost perfectly synchronized (ratio = 0.94) as both valleys (movement of shoulder centre and movement of abduction) are close to align.

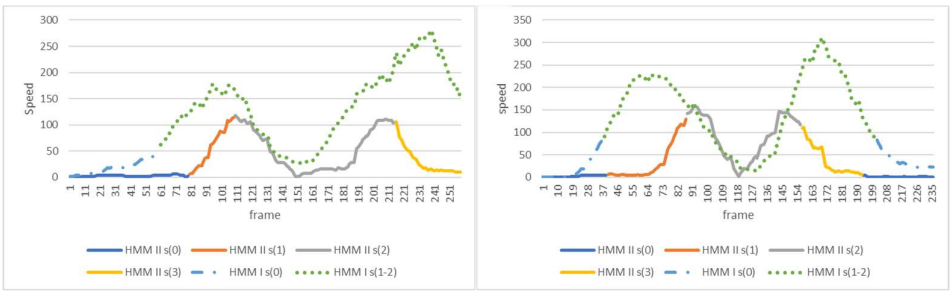


Figure 7.13 Speed path of the shoulder centre (discontinuous lines) and angular speed of the right hip (continuous lines), for two different trials of an exercise of hip abduction: trial 7 of subject 2 (left) and trial 2 of subject 1 (right). The linear speed of the shoulder centre is 1500 times magnified in this representation. The x-axis represents the number of frames, where every frame has a period of 1/60 sec. The y-axis represents the velocity in m/s for the shoulder movement and θ /s for the hip movement.

Table 7.9 Synchronicity between shoulder movement and symmetry of shoulder movement. Values close to 1 resemble perfect synchronicity or symmetry. For synchronicity, values <1 means the shoulders moved before the hip abduction and values >1 mean that the shoulder centre moved after the hip abduction. For symmetry, scores <1 means the shoulder centre reached the maximum amplitude faster than it returned to the beginning pose and the opposite for scores >1. For values >1 an inverse value (1/value) is given in parenthesis to be able to create an average synchronicity/symmetry value.

	Subject 1 synchronicity	Subject 2 synchronicity	Subject 3 synchronicity	Subject 1 symmetry	Subject 2 symmetry	Subject 3 symmetry
Trial 1	0.93	0.79	0.95	1.09 (0.92)	0.79	0.95
Trial 2	1.06 (0.94)	1.16 (0.86)	0.74	1.29 (0.78)	1.0	0.94
Trial 3	0.86	0.91	1.21 (0.82)	0.82	0.93	1.4 (0.71)
Trial 4	0.73	0.61	0.64	0.85	0.60	0.87
Trial 5	1.54 (0.65)	0.74	0.73	1.54 (0.65)	0.78	0.82
Trial 6	0.93	0.93	0.71	1.01 (0.99)	0.83	0.76
Trial 7	0.74	0.72	2.14 (0.46)	0.78	0.68	3.4 (0.29)
Trial 8	1.01 (0.99)	0.84	0.87	1.16 (0.86)	0.76	1.0
average	**0.85**	**0.8**	**0.74**	**0.83**	**0.79**	**0.79**

The created representation seems to capture to some extend the levels of synchronicity and compensation as perceived by the therapists. An example of a correct exercise can be seen in Figure 7.14 (left) where the symmetrical value (for the HMM I transition path) is 0.94. However, there seems to be extra reasoning involved in the ranking process of the therapists as low values in symmetry or synchronicity not always result into lower scores by the therapists. In case of execution 7 by subject 3, the subject was interrupted during the execution verbally by a therapist who had an extra comment. Due to this interruption, the last part of the movement was inhibited (low scores on both symmetry and synchronicity, see Figure 7.14 right) but not ranked as lesser. This shows that the therapist can cope with the interpretation of distraction within the subject and if it is justifiable to still mark the exercise as correct. As can be seen in Figure 7.14 (Right), the HMM predicts an earlier transition back to state 0 (rest state) than we can visually judge. The second speed hump has due to the interruption been classified as being part of the rest state.

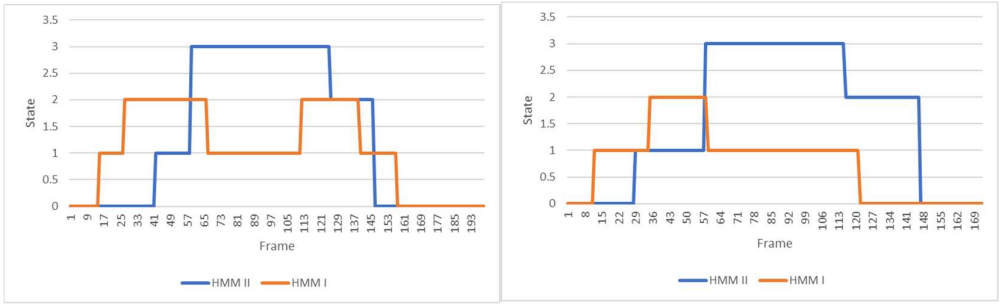


Figure 7.14 Left: exercise 2 performed by participant 2, where the HMM I and HMM II pattern is associated with a correct chronology of the movement. Right: Predicted state transitions for participant 3 trial 7 for both HMM I and HMM II where the HMM I pattern is missing a 'bump' (state transition). Speed on y axis and frame number on x axis (60 frames per second).

The low calculated score for exercise 4 (synchronicity: 0.61, symmetry: 0.60) performed by participant 2 does not match any of the therapists labelling. In Figure 7.15 the state segments can be seen of this exercise. The centre of HMM II s (3) and HMM I s (1) first and last transition do align resulting in a bad coordination score. This can be seen as an incorrect conclusion as the therapists did not label this exercise as containing lesser coordination and the zero crossing (HMM II) does seem to align with the HMM I middle s (1) segment and thus the way coordination should be calculated needs to be updated. Additionally, the lengths of HMM I s (2) are similar in length. This indicate that the structure of s (1) of HMM I does not influence the therapists rating. And that similarity within the states of higher speeds is most dominant in the assessment, which is imaginable as therapist are likewise more likely to detect deviations in conspicuous movement. Also, we can see the same effect in Figure 7.16 where execution 5 by participant 1 (synchronicity: 0.65, symmetry: 0.65) did not show up in the therapist's labelling. On the contrary a lower score of the therapists on exercise 2 performed by participant 2 shows in the state representation a slight dissimilarity in length of s (2) HMM I (Figure 7.17, left). In Figure 7.17 (right) an expected state transition, indicated with a red dot, to s (0) is not predicted by HMM I, resulting in a low symmetry score (0.64) but not a low therapist score.

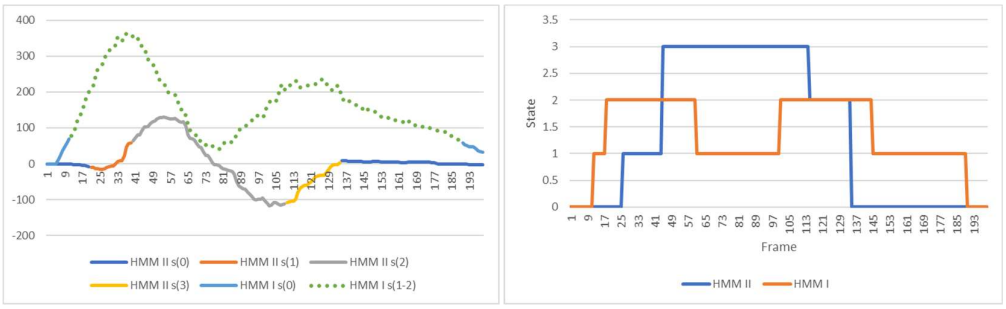


Figure 7.15 Exercise 4 performed by participant 2, where on the left the original signal is presented with colorized the different states and on the right the schematic state transition prediction of both HMM I and HMM II. Here the synchronicity and symmetry are both calculated to be around 0.6. Here this can be expressed as the exercise not being fluently. If we look at the state transition, mostly the last state is elongated showing the trunk moved longer than expected.

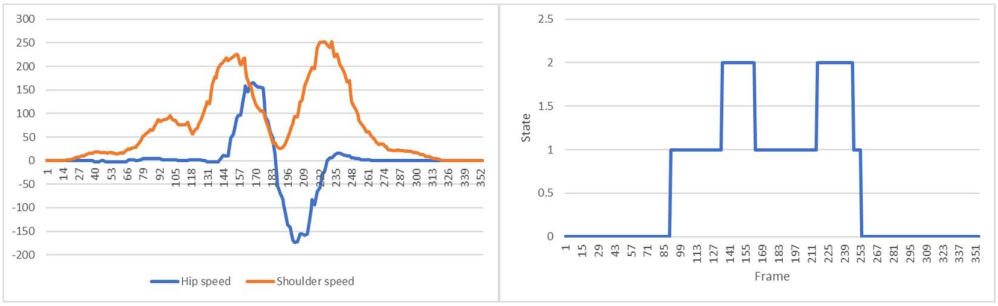


Figure 7.16 Exercise 5 performed by participant 1, where on the left the original signal is presented and on the right the schematic state transition prediction of both HMM I. Likewise as in the previous figure the exercise can be expressed as not being fluent as the initial s (1) is elongated. Again, the zero crossing of the Hip speed does seem to align with the valley of the shoulder speed indicating that repeatedly the low synchronicity measure (0.65) is incorrect as it does not align with the therapists labelling.

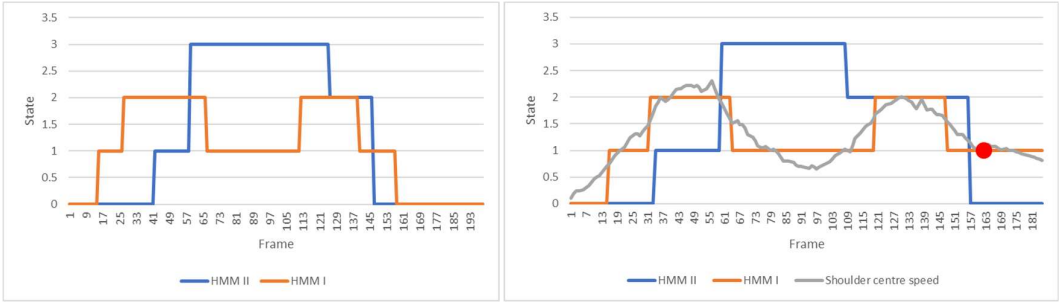


Figure 7.17 Left: exercise 2 performed by participant 2, where s (2) segments of HMM I shows to be of unequal length (accounting for lesser coordination quality as labelled by some of the therapists). Right: exercise 4 performed by participant 3 where an expected state transition (red dot) was not predicted by HMM I. This can be caused by the calculation of the actual speed value (this could be more accurately, e.g. diverse types of impact scales or smoothing functions).

One example where a therapist ranked an exercise as containing bad coordination is shown in Figure 7.18. As the results pointed out the synchronicity is 0.93 (1 is perfect). However, the aspect of asymmetry in speed (left and right) could have made the therapist decide to rank the coordination as not being perfect. An additional step that can be performed to estimate the synchronicity is applying curve fitting to the middle sections of the speed paths with the help of the state section selection. Then, calculating the shift between the valleys indicate the difference in timing between the signals.

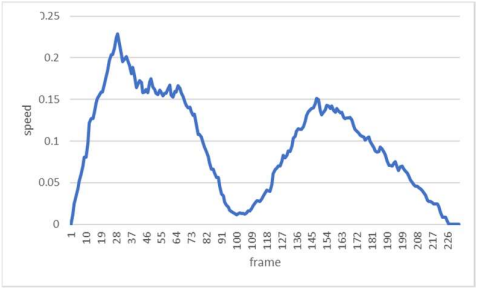


Figure 7.18 Speed path of shoulder centre for participant 2 trial 6 where there is dissimilarity visible in the range of speed. This compensation can be reviewed as a slight instability. Reviewing the video evidence, the subject was fluctuating slightly in the sagittal plane during the first phase of the movement (as this is an absolute value the area under the curve, speed dissimilarity, can differ with occurrence of instability).

7.4.5 Conclusion

Capturing a coordination value seems plausible with the usage of state transition paths of multiple trained HMMs. In this representation, the states with higher prototype speeds are dominant in the assessment of this coordination. The symmetry (compensation/) value should be transformed into a structure correctness measure so that it can provide more insight in were the exercise dissimilarity occurs. This calls for an approach in which the distributions of the states over time are learned so that states that are missing/added or elongated/shortened are directly visible as performance measurement. These measures will provide insight in the stability and flexibility factor of a performed exercise. Additional work should acknowledge the inclusion of the total displacement of the trunk during the movement and a symmetric value for the distribution of the displacement. Where total displacements can aid in detecting the allowed movement where it is not considered compensation and symmetry in displacement can tell something about the smoothness of the movement. This means that a coordinate system should be created in which displacement and motion can be expressed in relative and not absolute terms. Furthermore, the HMM predictions sporadically contain noise as can be seen in Figure 7.19 (rapid state changes). With the mentioned learned state duration distributions, these misclassifications could be identified prior and corrected for while computing a synchronicity value. Or instead the exercise could be discarded of and don't get evaluated.

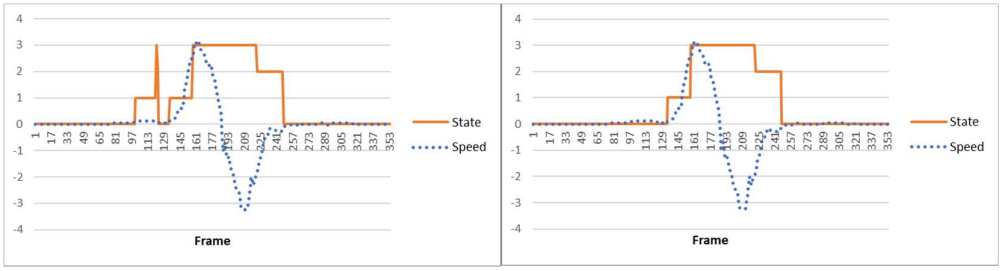


Figure 7.19 Left: example of noisy state prediction (HMM II) in exercise 8 executed by participant 3 on the right the corrected version.

7.5 Experiment IV – Patient/Healthy subject comparison

This concluding section regards an experiment where the developed method as described in the previous section is applied on gathered data from real patients. Here an analysis of the classification results and ontological differences detected will be discussed. The goal is to see how these patients need to be conditioned to be able to perform the suggested exercises as good as possible and whether the previous explored method needs to be drastically or minorly altered. This experiment took place in a public hospital (*Hospital Especializado de Atencion Integral del Adulto Mayors*) in Quito. Figure 7.20 shows the therapeutic practice where the patients were recorded.

Can the method be projected onto real patients?

Figure 7.20 experimental setup, where the person recording is not facing the patient to be less emphatically present in the process.

7.5.1 Introduction

The HMM approach developed in the previous section is applied here on the data gathered from 2 elderly patients that are within the early recovery phase after hip replacement surgery. The conditioning of the patients is monitored as well as simultaneously the effect that it causes on the data. Conditioning here is the patient representation (clothing etc.), therapist positioning, therapist guidance during the exercise, and feedback provided. Furthermore, the therapist labelling is compared with calculated forward probabilities and movement state sequences to synthesise more restrictions on what is presumed as correct/incorrect. Some suggestions are provided on the 'dress code' during interaction with the 3D motion capture device as some data was corrupted due to partial body occlusion.

7.5.2 Protocol

Two patients (74, and 77 years of age) participated in this experiment. These patients were attending their regular session of physio therapy for their rehabilitation. Both patients had undergone hip replacement surgery at their right hip. This surgery has taken place 1,5 and 3 weeks prior to this experiment (early stage recovery). They were informed on the gathering of their exercise data and its later use for scientific research purposes. With consent, video of the whole session where created as well as depth images to later be able to extract the skeleton and apply the input transformation into the desired feature representation. The patients were asked to perform 5 different exercises, each exercise was repeated 8 times (total recordings: 80). The executed exercises were: Hip abduction, Hip extension, Hip flexion (standing), Hip flexion sitting, and a three-step exercise (step forward followed by step to the side and later step backwards). The recording is only stopped during exercise change so that the therapist can guide the start of every sequential execution within the session. The person recording is facing another direction, to minimize the sensation of being watched (Figure 7.20). The patients are the clients of the therapist that guided the entire process of patient conditioning. A second therapist was present at the same time. Both preformed the assessment similar to that of the previous experiment.

The trained HMM's of the previous experiment (HMM I and HMM II, Table 7.7) are used here to classify the movements of one patient and compare the results to a healthy participant (participant 1 of previous experiment). In addition, their differences are briefly visualized. The therapist labelling and recordings are then used to determine if the detection of compensation is correctly obtained by the HMM. The videos here aid in the actual error that occurred.

7.5.3 Results

Conditioning – As these patients are relatively old, the therapists performed the exercises a couple of times with a clear slow description of the action itself. The first patient had problems with memorizing the order of the three-step exercise, showing that the cognitive abilities play an import role in the selection of suitable exercises in the rehabilitation plan. The patients both wore baggy clothing that did not afflict with the therapist's assessment. This did however, in some executions, led the tracking of the skeleton to be corrupted and created data that was unsuitable for analyses. The second patient needed to perform the exercises with a walking chair as there were still stability issues. This chair did not impact the skeleton extraction on a same level as the clothing did. Figure 7.21 shows the appearance of the patients. During the exercises, the time that the patients where in the most extreme pose (full abduction range) the therapist did communicate in some executions to drop to the initial pose. Here the patient needed to be nudged to finish the exercise. Minor feedback was provided after every execution, mostly the feedback consisted out of the possibility to reduce compensatory behaviours (stand up straight and let the force be applied form the hip) in case of the hip abduction.

Figure 7.21 The recovering patients as executing Hip abduction during the recording session

Classification comparison – An analysis is performed on the executions of patient 1 regarding hip abduction. This is compared with the executions from participant 1 which will here be referred to as Heathy subject. As can be seen in Table 7.10, according to both therapist's execution 4 and 5 are of lesser quality (contains compensation to some extend). For the labelling of all the exercises see Appendix C. For exercise 4 this Cleary shows (Figure 7.22) in the classification of HMM I (trunk) and in the labelling this exercise received the lowest overall score as well.

In Figure 7.22 the classifications of both patient 1 and the heathy subject can be seen. The average probability of correct labelled exercises here is 1261 for the patient vs. 1474 for the healthy subject which indicates an average better execution by the healthy subject. what is noticeable is that the classification of the patient execution gradually increases over time (not considering the execution 4 that is considered to contain compensation). This means that the patient did increase its performance over time. On the contrary the same effect does not seem to appear for the healthy subject. This subject did perform the exercise better over time without any feedback until exercise 6. However, the small brake within the recording session (therapist discussion about possible feedback provision) between execution 5 and 6 for this subject can explain the classification drop to some extend (loss of flow). The errors occurring in trial 4 and 5 will be discussed in the section Error analyses. The threshold can be defined in terms of relative performance compared to previous executions. For example: the prediction should at least be higher than 70 percent of the averaged previous 3 executions.

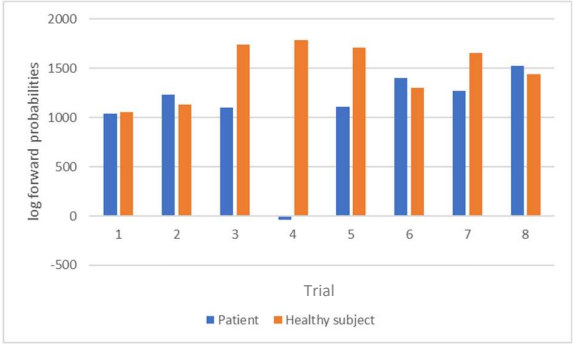


Figure 7.22 HMM I (trunk) classification of 8 consecutive executions of patient 1 (blue) and a healthy subject (orange).

Table 7.10 Therapist labelling of hip abduction for patient 1

	Therapist 1			***Therapist 2***		
	excellent	*good*	*bad*	*excellent*	*good*	*bad*
ROM		1 2 3 4 5 6 7 8		7 8	1 2 3 4 5 6	
Coordination	2 3 4 5 6 8	7		1 2 3 4 5 6 7 8		
Compensation	3 4 5 6 7 8	4 5		1 2 3 4 5 6 7 8	5	4
Force	3 4 5 6 7 8			1 2 3 4 5 6 7 8		

The overall appearance of certain speeds in the trunk movement is shown in Figure 7.23 for patient 1 and the healthy subject. A histogram is created for bins that contain exercise values of trunk speed. The total amount of data points in these histogram representations are 2201 for the patient and 2382 for the healthy subject (both for the total correct executions). This representation shows that speeds in the executions of the healthy subject stay lower on average that for the patient. For the healthy patient is seems that it is rather unlikely that speeds exceed values over 0.15 (m/s) where for patients the distribution for values over 0.15 (m/s) are comparable in appearance. This measure could aid in detecting whether the patient is in an early or later state of recovery as it shows a periodic variance that can be considered as a consistency measurement.

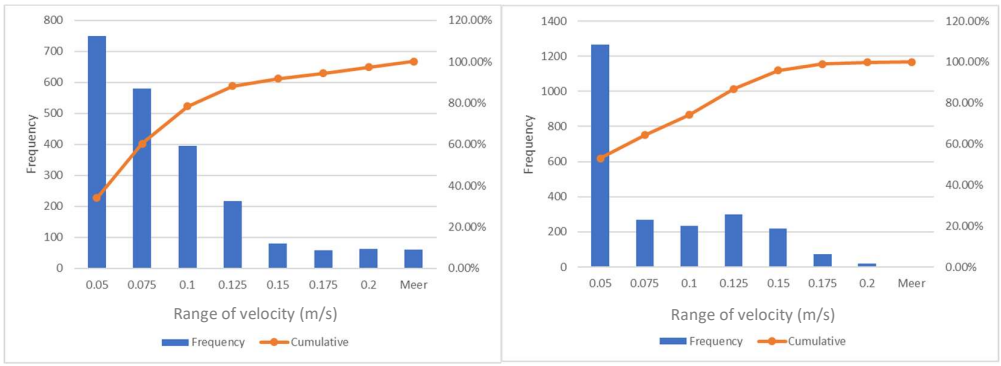


Figure 7.23 Histogram representation of trunk speed in hip abduction, left for patient 1 and right for the healthy subject

Error analyses – Further detail on the compensation in execution 4 and 5 is provided in this section. Analysing the video of exercise 4 shows a clear compensation. This compensation was an uncontrolled deviation of the hip (centre) moving to the left at the end of the movement. This is shown in Figure 7.24 where the yellow line shows the speed of the hip centre in this compensatory movement (left) compared to a correct execution (right) by the patient. This compensation can be revered to adduction and occurred in the left hip. This suggest that the stability of the patient during the execution was not yet of the quality required to perform the exercise correctly as the mussels in the non-target hip could not support the body wiles returning to the initial posture.

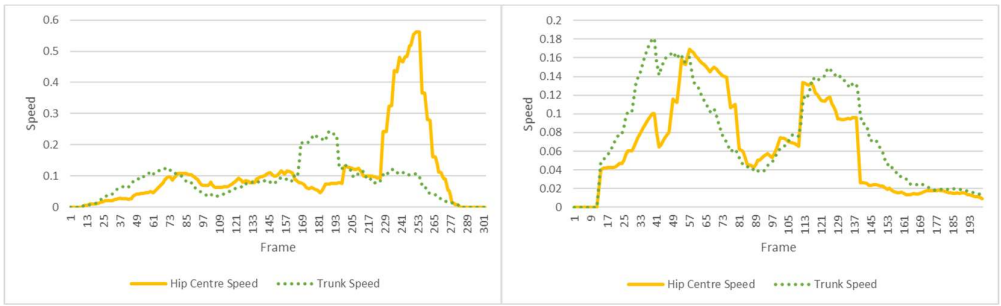


Figure 7.24 trunk movement of patient 1, left execution 4 where in the end of the movement a compensation occurred and on the right execution 3 that was identified as correct.

Execution 5 was not classified inherently different to correct executions. In Figure 7.25 (left) the trunk movement of this execution is show in comparison to a correct execution. An additional 'bump' can be seen in the patient's signal. the video shows that whiles approaching the maximal range within the execution the patient lost balance resulting in a minor 'freefall'. However, the patient was able to correct this in time and successfully terminated the exercise. This imbalance can be clearly seen when predicting the state transitions path (Figure 7.25, right). A correct execution as demonstrated in the previous experiment has 2 'bumps' where here the mid-section has an additional state transition into state 2. This shows that the state transition predictions are an indicator of com-

pensation, perhaps even more so than probabilities as it provides additional restrictions on transition order. Figure 7.26 shows additionally that the trained HMM II (Table 7.7) can segment the patient's target into the state sequence in the same order as for the heathy subject (transition of 0, 1, 2, 3, 0 example in Figure 7.15).

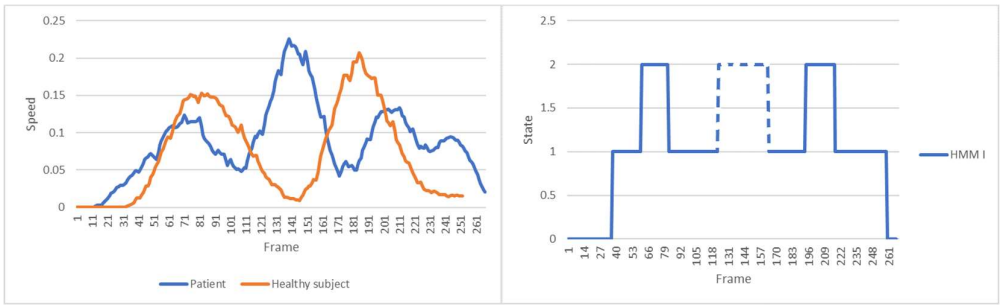


Figure 7.25 Left execution 5 by patient 1 compared with a correct execution of the healthy subject (Trunk movement), right the predicted state transitions (HMM I) of execution 5 by patient 1 that reveals the additional state transition (dashed) caused by compensatory behaviour.

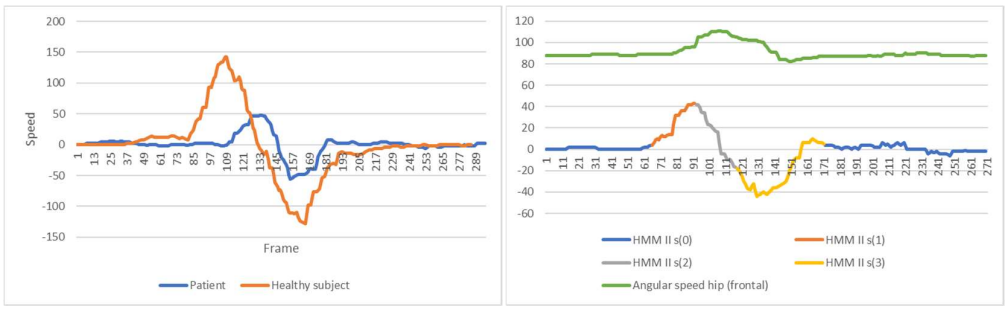


Figure 7.26 Left a comparison of hip speed in the frontal plane for the patient and healthy subject where shapes are similar. Right the segmentation (of the patient execution) as predicted by the earlier developed HMM II.

7.5.4 Conclusion

In this experiment, we can conclude that patient's initial conditions (clothing) are important in the accurate detection of the skeleton. In addition, some differences where noticed comparing a patient to a healthy subject. These differences where both visible in the average classification and averaged speed distributions. Where average classification is higher for the heathy subject and the speed distribution has a lower variance. The classification values are however, not substantially higher. This means that initial patient status (classification values) should be considered because the outcome of correct trials can be classified lower than that of a healthy subject. Classification with only the use of forward probabilities seems insufficient to capture varying types of compensation but state transition sequences seem to be able to uncover certain of these undetected compensations. A possibility to detect this sequence mismatch is visualized in Figure 7.27. The target HMM II (Table 7.7) shows to provide a well working segmentation.

Suggested is an additional requirement regarding the execution for exercises. The use of tight clothes will be preferred and advised whiles patients preform exercises. This can be in the form of sportswear to ensure the flexibility whiles performing an exercise. State transition sequences seem to clearly show additional important clues about the occurrence of compensation. Therefore, as earlier mentioned time restricted state transition probabilities should be introduced into further work were this can be transformed into a forward probability measure in the same fashion as performed in this work. Average classification values and variance can be used to estimate patient progress with the use of multiple consecutive executions.

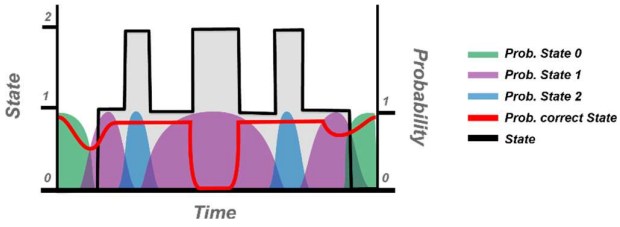


Figure 7.27 Versioned detection of being in an unlikely state (red line). Where the expected current state is visualized as a distribution over time and the example transition path of execution 5 by patient 1 is used as the example. Then, clearly the undesired state clearly shows (middle).

Conceptualization

8 Assessment methodology

With every experiment conducted the vision on how to construct the assessment is further shaped. Redefining features and structure of the HMM implementation led to a methodology that will be discussed in this chapter. The process of feature creation, HMM training and assessment blueprint will be explained. Usage of these models and its anticipated flexibility, partially shapes the structural constituents of the platform. Therefore, besides the assessment methodology, proposed integration is discussed on a level of data creation, storage and feedback generation.

8.1 Model Ontology

For the feature representation as proposed, additional data is required. Not only the joint positions will create the motion representation, in addition the depth data and joint orientation data will be included in the raw data capturing. The joint orientation is a 4d vector (quaternion) containing the coefficients to calculate pitch, yaw and roll between two sets of coordinate systems (in this case the joint specific orientations). With these rotations, mostly the detection of harmful movement will be enabled. This depth data is needed to create the personal coordinate system so that motion can be expressed in terms of relative movement in frontal, sagittal and axial direction. The proposed total feature representation can be seen in Figure 8.1.

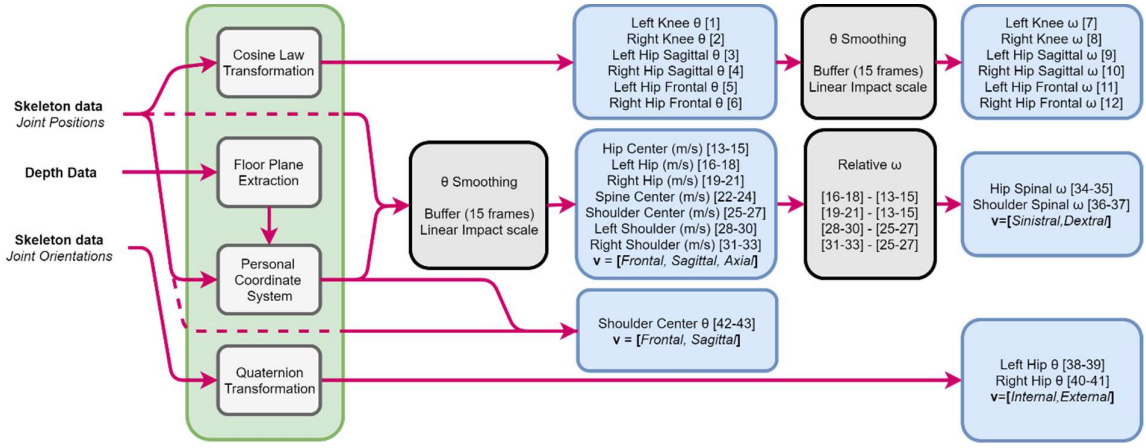


Figure 8.1 Proposed feature representation. In green the required processing steps, grey indicates minor transformations and in blue de sets of features. The arrows indicate the direction of the data flow.

8.1.1 Angles

The transformation from joint positions to angles is one of the straight forward tasks that have been shown to be computed rapidly utilizing the cosine law (Equation 8.1). First the absolute lengths between joints are calculated (Pythagoras theorem) where one angle is extracted as feature. Here the first features are created to represent the relative pose of the subject. these are features 1-6 as can be seen in Figure 8.1. Additional vectors need to be created in a similar fashion as described in the next section. This vector enables the correct representation of hip rotation in the sagittal plane (Figure 8.2).

$$Cos\ (A) = \frac{-a^2 + b^2 + c^2}{2bc}$$

Equation 8.1 Cosine Law

8.1.2 Floor Plane

The floor plane plays an important role in creating a personal coordinate system (PCoS). The need in this scenario is to find vector perpendicular to the floor that crosses the skeletons reference point (joint 0, hip centre). With this approach, floor orientation is not considered (is the floor truly horizontal e.g.) but this could if necessary be an additional parameter. With the earlier mentioned VFH (section 4.6) a part within the depth image can be scanned in the neighbourhood of the feed joints. Here the dominant surface normal within the histogram representation will be used in its opposing form to project a scalar onto the floor. This creates the axis of rotation on which the PCoS can be constructed.

8.1.3 Personal Coordinate System

With the floor plane orientation, the PCoS can be constructed. The additional data required in this step are the left and right hip joint position data. The vector that crosses both these points can be translated onto the reference point. Now at this point two vectors meet at the reference point and enable the third direction to be extracted by finding the perpendicular of these directions. This can be done by using the cross-product rule (Equation 8.2).

$$A^{\rightarrow} \times B^{\rightarrow} = [a_2 b_3 - a_3 b_2,\ a_3 b_1 - a_1 b_3,\ a_1 b_2 - a_2 b_1]$$

Equation 8.2 Cross-product rule

Using the newly found orientations, the absolute coordinates per joint can be translated into personal coordinates where the origin is the floor crossing with the earlier mentioned scalar (see Figure 8.2). Note that the PCoS is rotation invariant so that movements can be expressed in terms of front/back, left/right and up and down. The speeds in the 3 directions found are differences in orientation between consecutive frames. Here the speed is calculated as an averaged sum of the previous 15 frames (±0,25 sec) where there is a linear correlation between recentity and impact on the calculated value. The rotational values (features 34-37) are still calculated using the absolute values.

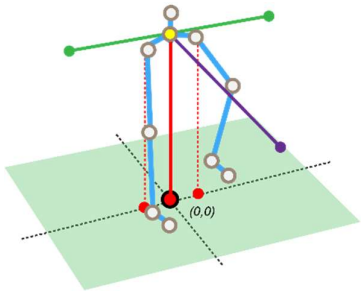


Figure 8.2 vectors creating the personal coordinate system with the centre floor point being the new origin. The dotted lines indicate the addition vectors that are required to calculate the hip movement in sagittal direction.

8.1.4 Rotations

The rotations that we need to capture are those that resemble external and internal rotation of the upper leg. With the joint orientations, quaternion values can be transformed into different rotations (x, y, z of parent axis). These values are initially used in animations to recreate the rotation of limbs easily. A hierarchical bone structure and a coordinate system per joint specific, enables the rotations to be extracted from the quaternion [93].

8.1.5 HMMs

The features are input in learning the parameters on the level so that semantics comply with therapist's analysis of an exercise. Most coarse types of movements (with this semantic similarity) are incorporated into the HMM representation, where the restriction of the Skeletonization algorithm are also considered (no trunk bending possible). This means that the earlier developed multi HMM assessment is advanced into a 10-fold semantic manifestation that is represented in Figure 8.3. For each of these HMMs, 2 measures will advance into the assessment blueprint. The forward probabilities are used with a sliding window to indicate the likelihood of each phenomenon within the movement. This pinpoint the location of an error, if this occurs in one of the HMM elements. Secondly the state transition sequence is passed on to be analysed on symmetrical values and coordination between the different HMM elements. When the forward pass provides a low likelihood, the corresponding state mean values (those containing directional speed) are compared in an absolute fashion to determine the direction of error. To note is that, prior to assessment, the correct pose needs to be adapted by the user, which likewise can be determined with the developed features and learned threshold values (HMMs initial states distribution likelihood).

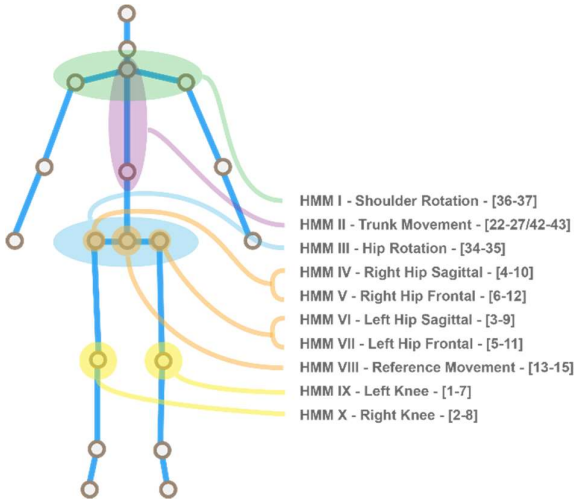


Figure 8.3 The 10 HMMs that are trained on specific features to represent movement of specific meaning. In brackets the used features per HMM.

8.1.6 State analyses

Besides the forward probability mapping, the duration of states can be modelled into probability density functions to provide a template on which state time durations can be expressed as a sequence of likelihoods. Hence, it could occur that multiple state sequences are correct and should therefore be stored so that the structure of a to be classified execution can be assigned adequately. In addition, the correlation between the state lengths will be extracted (using curve fitting, Figure 8.4) to exclude misjudgement of states that are less dependent on states prior or posterior. One of these states could be the reached maximum range, where some subjects hold the pose for a longer period. In this way, this extreme pose condition can be automatically detected. Whiles in this state and exceeding the state duration within a probabilistic interval, feedback can be generated to let the user to 'let go' as this is not a true error but merely a style of execution. The last and first states are clipped to an equal length for all executions before the duration distributions will be calculated.

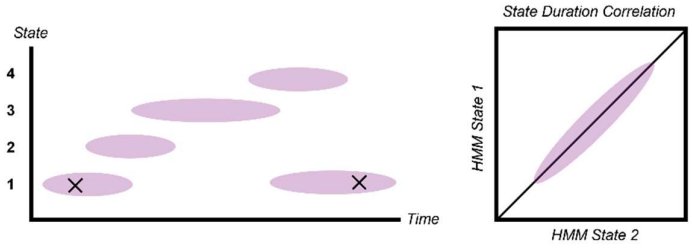


Figure 8.4 left a representation of state duration distributions where the PDF here is negatively (left) skewed. On the right, a representation of the possible fitted curve that correlates state lengths.

States that do depend in length on each other can be corrected depending on the context. For example: if a subject would perform an exercise in a relatively fast paste, but one of the state durations is that of a more likely occurrence (state duration is mean value), the overall judgement will state that the exercise was executed faster than regular but there was a part where the paste of execution was slower. Initial and more important terminal state combinations of the 10 HMMs can be extracted to automatically clip an exercise into the right size (marked as x in Figure 8.4), optimizing storage and classification of the exercise.

The state transitions will also provide a value for the coordination between the target HMM's transition path and an arbitrary other HMM's transition path. Instead of creating one value for coordination, each state (in a learned sequence) obtains its own coordination value. The States durations plays an important role as the expectancy whiles time progresses of being in a state (one that is of another HMMs prediction) changes in the distribution of the correct executions data in a well-coordinated fashion. In short, a likelihood will be calculated; given a current state duration (Figure 8.5), what is the probability that the predicted state of another HMM is at the same time in a given state. A problem that could occur is that when there is a state transition mismatch (i.e. a compensation) the coordination values alter rendering it useless. There are multiple options to work around this, like reclassifying the sequence with a restricted HMM (outputs only one type of sequence, not a fully connected Bayesian network).

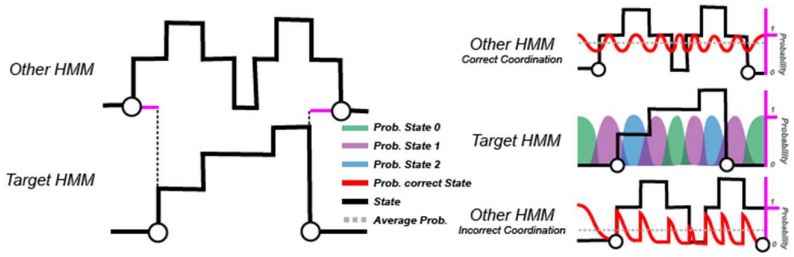


Figure 8.5 Left, two state transition paths and the simple coordination measure as created in experiment III.
Right, the proposed coordination measure with for each state transition within the prediction of the target HMM an associated probability value.

8.1.7 Assessment blueprint

Here the total set of assessment measurements with their semantic description is presented (Table 8.1). 7 different assessment values are incorporated in the evaluation of the correctness per HMM prediction. These measures are created in such a way that it carries therapeutic relevance. The pipeline of assessment is choses to first asses the most clearly detectible faults. At first a question will be asked whether the movement has compensation or if the movement was executed to rigidly. Then the speed per state is assessed to detect slowness or to speedy executions. With the error correlation of the state durations the overall fluency is measured, and the general speed indication can be determined. Then ROM and pose holding is calculated and last the synchronicity. In case of state sequence mismatch in step 2, the additional transitions will not be considered and if there are any missing transitions, these transitions will be estimated trough state duration/correlation expectancy. Altered transitions (e.g. expectancy of a transition matches the altered transition) will be incorporated into this pipeline step.

Table 8.1 Assesment blueprint with individual evaluations that (I) are applicable to the target HMM and (II) the 9 additional (other) HMMs.

Pipeline Step	Assessment Values	*Other HMMs*	*Target HMM*
1	Forward Probabilities*	Unlikely Translation/ Rotation (compensation)	≡
2	State sequence mismatch (Probabilistic localization)	To rigid (missing states) or compensation (additional transitions) > counting transitions and a state speed prototype speed comparison	≡
3	Direction of faulty movement	Pinpoint the direction: left/right, front/back, up/down of the fault >>	≡
4	State duration probability	Correctness of speed per state	≡
5	State correlation error	Fluency of movement	≡
6	Maximum deviation State	-	Range and duration of the target pose
7	State transitions	Synchronicity of Translation/ Rotation movement	≡

8.2 Model Integration

The sets of algorithms that create the assessment blueprint needs to be constructed on available data. Here to implement these sets of algorithms, the dependency leads to several suggested modules to ensure that the therapist can act autonomously from any data scientist. The structure of the platform will be developed as suggested in Figure 8.6. Here different transformations of the data are saved to review the data on different levels. Visually projecting raw data (*motion capture module*) onto an avatar can be used to analyse an execution that is recorded by the therapist him/herself. The Kinect's data of the user will be transmitted via the internet (See Appendix E) into the database. This analysis can be saved as al labelled version coupled with the suggested feature representation. These sets can then be used to train the models (*Model training module*). When such a model is trained, the model can become an active part of the platform. Data of performed exercises will, in real time flow through the assessment. This creates classified data according to the blueprint. This data needs to be interpreted and translated to feedback to be useful in the patient interaction. This interpretation is done by the *progress & plan module* where the estimated point of progress compared to the quality of the execution is the establisher of the eventual feedback provided. This feedback could additional be saved to analyse the effect per type of feedback. Simultaneously the therapist can keep track of the progress in this module. In the overview, different data streams can be seen where in green the live interaction data flow of the patient is shown and in red processes that can be executed autonomously from the live patient interaction. The raw version of classified data with ambituses classification values in any of the 6 assessment measures will we highlighted in the mocap module as 'to be labelled'. In the next chapter, the different prospective modules are presented in more detail.

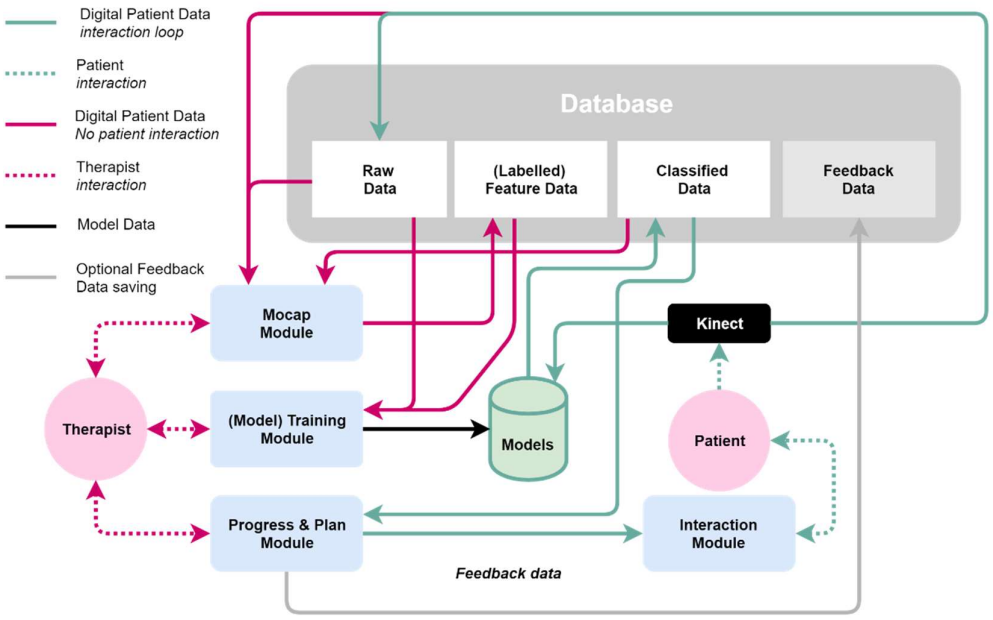


Figure 8.6 Proposed integration of the assessment model into the platform with the required prospective models to ensure control of the process by the therapist. In blue the necessary models, in grey the data types that needs to be stored and in pink the users that need to interact with these different modules.

9 Prospective Modules

The prospective modules are based on the presented method of assessment (previous chapter). These modules can be operated separately from each other. An architecture is implied that enables these modules to adopt multiple assessment methods. In this context, the assessment methodology as presented will be revered to as method ePHoRt.

9.1 Mocap App

As in the processes of the automation of therapeutic feedback generation, the first important step is to record and being able to label the data. Therefore, the mocap app is suggested (Figure 9.1). In this application, the therapist can record exercises, label the exercises in the same fashion as performed in the experiments (ROM, Compensation, Coordination) and replay any data (video and projections on avatar). When logged in, the therapist can select a recording session or create a new recording session. Such a session can be specified with a name and will receive a timestamp of the creation date. This repository will have its sub repository of exercise types (again created by the therapist). Here the target hip can be selected to later ensure that mirrored recordings are corrected for. This mirroring occurs when recordings are performed on a subject that execute exercises with the left hip instead of the right (or for that matter any other type of exercise). The connected device can be chosen and will enable the correct formatting of the recording data. This leaves flexibility to which devise will be used in the process and creates a tag per datafile of the used device (of later use when feature representations need to be created). Every recording can be quickly replayed and discarded/saved, as preferred. The labelling can be executed by selected a window (grey and can be dragged on left and right side to appropriate size) and click on a joint where an error occurred. The avatar as showed in Figure 9.2 can be inspected from multiple perspectives by dragging around this body. Here the therapist can later review an exercise to identify if any error occurred (human check-up). When clicking on a joint a form for the error appears where ROM Compensation and Coordination need to be assigned with – (noting), bad, good or excellent. The type of form that appears depends on the type of assessment method that is choses in the labelling preference. Videos as well as joint representations can be played back.

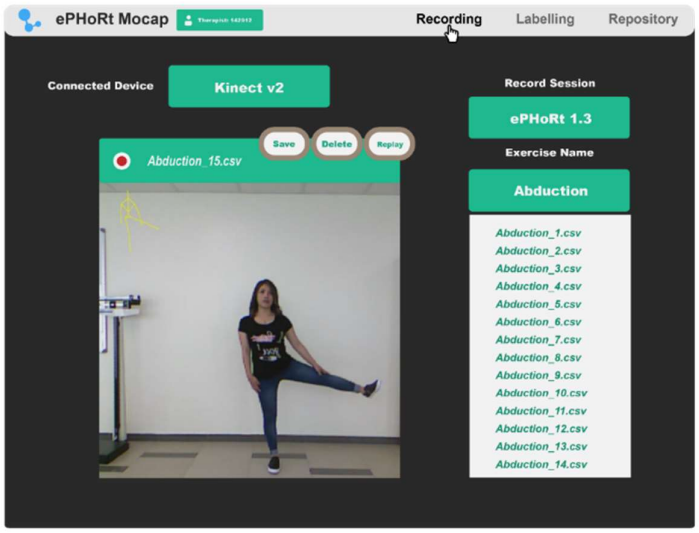


Figure 9.1 Example of the mocap module where the recording interface is visualized (device selection, recording session selection, list with exercises in current exercise recordings, video, skeleton and replay modes).

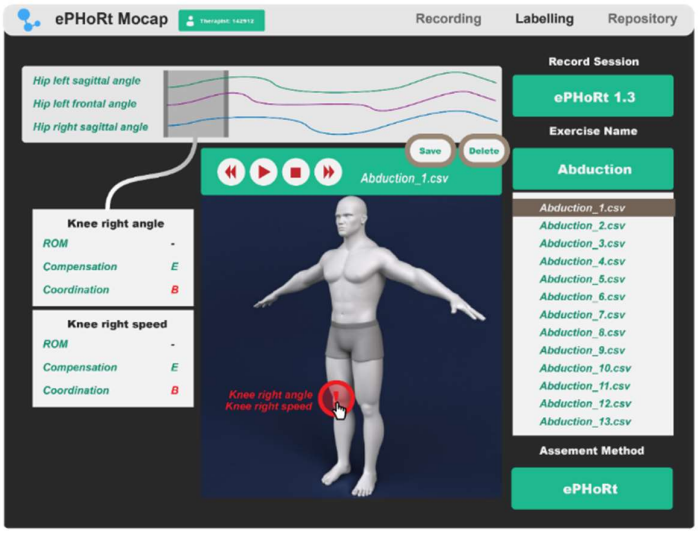


Figure 9.2 The labelling interface shows the label window (light grey between grey bars, can be dragged too), avatar with location of error, error forms of the location, the selected assessment method and as in the recording interface navigation within different recording sessions. (video/ avatar modus)

9.2 Model Trainer

The model training module (Figure 9.3) will operate the training of HMMs that will be mostly hidden from the therapist. The collected and labelled data by the mocap app can be loaded into this application to created accordingly the correct models to asses an exercise of the tagged exercise class. Firstly, the therapist will have the freedom to adjust their available models by updating the data input that is used in training and can create new exercise models. The labelled data here will be used to create a most accurate representation of a correct execution (by not training on this data, parts of data). In a later state, the labels can also be used in comparing classifications accordingly to further align the calculation of the assessment as described in chapter 8. Suggested is to use labelled data or in case of unlabelled data the assurance that the data is of a correct execution. The assessment method can be selected ones again to create the correct set of features (and blueprint matrix), addition the target value can be selected. When hitting the save button the model will be trained, and the therapists will be informed on termination of this process. State selection here is automated (with the use of BIC).

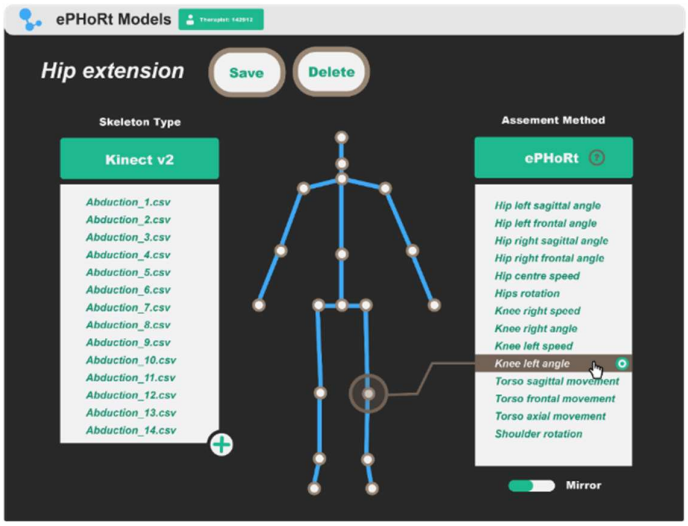


Figure 9.3 model training module where data can be selected (in this example first exercise hip extension is selected, left top corner), target joints can be selected, and the type of assessment can be assigned to create the right feature representation of the selected data.

9.3 Progress module

In this module, the therapist can track the progress of the patient (Figure 9.5). First the patient that needs to be reviewed should be selected (left top corner). In addition, a time frame can be selected to zoom in or out on the progress (in the example 30 days are selected). Per different exercises (that the therapist selected for the rehabilitation program) this progress can be reviewed on various aspects. These aspects will be briefly discussed here. First the amount of activity can be monitored (e.g. how often did this patient perform the exercise). Per day, notifications show up if any harmful situations occur so that the therapist could contact the patient at the appropriate time and could review this exercise by getting redirected to the exercise visualization part of the mocap app. Sliding through the days provides more insights on how the performance was on a particular day. The raw extensive classification blueprint is not shown here. Instead the feedback generator (another part of this application), that balances this assessment input with the current focus and progress of the rehabilitation reflects the feedback that is ultimately also provided to the patient. In this case (Figure 9.5) the feedback translates to short, trunk abduction or good. Per classification the goal target in terms of ROM or classification (or other) result is split up in an average difference to the ultimate expected target value (e.g. as good classification results as a healthy subject) and the deviation of the executions compared to the deviation of a healthy subject's executions (consistency check). In this application, not only the classification is stored and reviewable. In addition, patient initials (Figure 3.5) are stored as those are important to estimate the recovery time and expected progress. The initials will also carry focus points (in time, Figure 9.4) that will add weight as a maze to the assessment matrix.

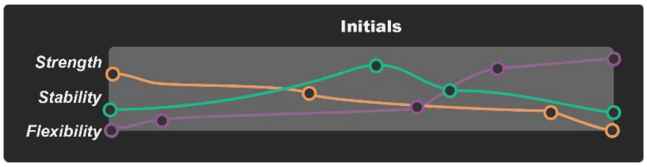


Figure 9.4 Part of the initial planning of the rehabilitation with the sift of focus over time, from strength to flexibility (dynamically adjustable).

At the start of the rehabilitation program a plan needs to be created with the set of exercises that the patient will need to execute. The evolution of frequency of execution can be defined in the plan as well. Or in case of an adjustment in the rehabilitation plan exercises can be deactivated or activated. This progress module can combine the estimation of the progress and current execution classification into a feedback stream. Progress here can be measured as the difference in classification results or ROM values where these are compared to a pre-set target in terms of mean value and deviation of a set of executions.

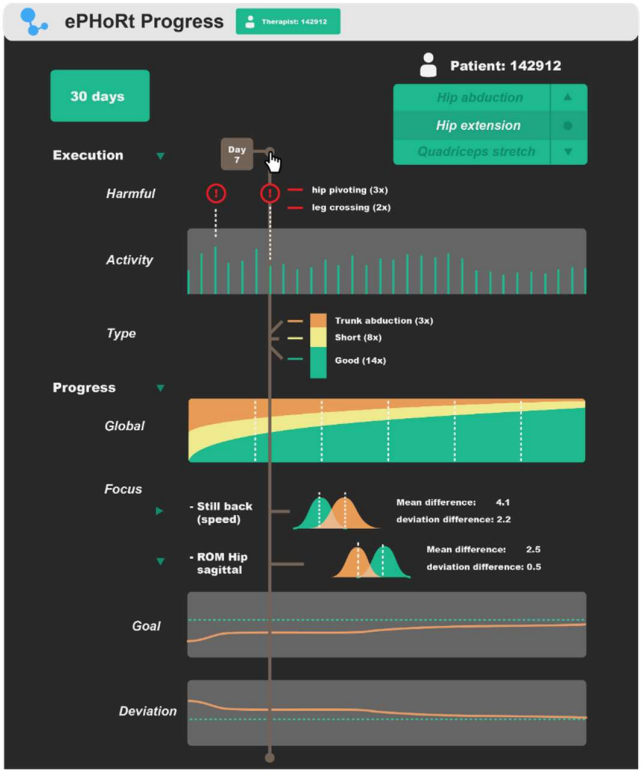


Figure 9.5 progress module where activity can be tracked per exercise type and global/local progress can be reviewed.

Conclusion

10 Discussion

Assessing quality of therapeutic significance has shown to be highly dynamic. Due to this the processes of analysing misclassifications have been essential in this work. Transcending from a controlled and acted exercise execution to patient execution, brought the methodology to a state of being implementable into the platform. In conducting the experiments, the focus has shifted from creating classifications from a top-down data driven perspective to a knowledge based bottom-up approach (qualitative). The first two experiments, where DTW and HMMs are utilized, show that it is possible with high accuracy to distinguish (I) good from bad executions and (II) different types of compensations if data is gathered when classes are in balance. The applicability of this approach however could be unsound as the gathered data is expressed in a way where the subjects were asked to substitute into a role. This acting restricts the classification to clear class division and does not consider the abundance in possible faults and combinations of compensations. But with the acted executions, the exaggerated faults where straightforward to analyse. With every new experiment, unforeseen articulations in several types of errors, extended and refined the feature set. In essence, acting assisted in a smoother analysis and defining clear improvements. The use of HMMs over DTW is stimulated by the fact that the performance of the algorithm (DTW) is very sensitive to the database size [80] (in contrast to HMMs). As the incentive of project EPHoRt is to increase the database on the fly, this conclusion substantiates the use of HMMs. Also, the use of HMMs provide insight in the ontological structure (semantically decisive). Considering that HMMs are generative models, it is possible to find out how categories differ from each other (based on the distribution differences).

In the later experiments, the subtler differences in the executions showed that therapists are not always consistent with their classifications (interpersonal). The human, even being professional, had its own personality (biases) within classifying an exercise. This should be considered as in the end the therapist in this use case, will still be in control of the patient's progress (Individual therapist augmentation). To note is that in some cases, where the therapists were divided at two tables, there was more consistency per table than in totality. This was caused by occurrence of small discussions between the therapists per table. For a more consistent assessment methodology development, the consistency within the therapist's judgments would be desirable. The gathered video images can be used to reach a large group of therapists that can label individually (mechanical Turk). With a larger share of labelling, judgment distributions can aid in finding the most general assessment within an exercise.

Creating speed representations of the angular change is done by buffers that has a linear impact. This is not yet optimal as speed should be derived locally. Different buffering techniques could be explored to approximate the true speed representation. As suggested in 5.3, the use of curve fitting techniques could minimize the need for data storage and in terms can make it possible to be derived locally. Bezier curves can describe a time series into predefined constituting motion paths (***pathobs***). Here a sequence per motion is build out of consecutive segments (Figure 10.1) that have specific start and end specifications that are defined by the functioning of muscle tissue in motion (Figure 5.2). This representation is powerful as it vastly shows the structure (how many parts, what kind of parts) and creates the possibility to be adjusted segment specific. The adjustment can be used in animations and better understanding of the impact of 'The solution's behaviour' with the initial conditions, by letting altered versions to be judged by a therapist. Switching linear dynamical systems (SLDS) propose a similar approach as it tries to model data as a linear projection of a low-dimensional latent state. This SLDS algorithm have shown to outperforms DTW in terms of correlation with a clinical evaluation [68].

The feature representation here is based on the skeletonized image that is supplementary to the motion capture devise used in the study. This representation should work as a high-level language intermate to therapist visual understanding and varieties of Skeletonization or other body featurization algorithms. The personal coordinate system as suggested in 8.1 can regularize the different featurizations into angular change of every joint and displacement of the subject within the sensorial space. Without the ability to distinguish between different planar movements (as for trunk movement in the experiments), compensation behaviour could stay undetected or classification errors could be corrected for in an unsound way. This intermediate language is of great importance in the analyses of any type of error and should therefore be further developed and standardized.

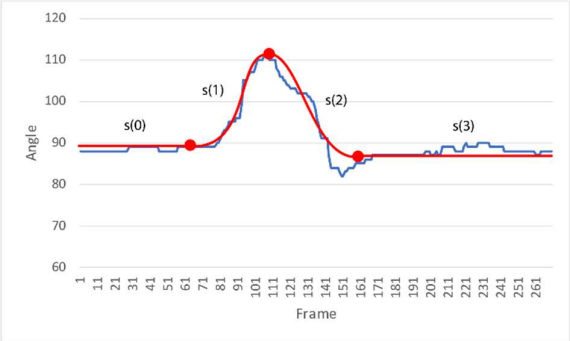


Figure 10.1 Example of a signal that is split up in fitted segments that have predefined start and end conditions (0 gradient).

As mentioned in 8.1 additional features can aid in the detection of harmful movements. This fact has been underexposed but will be essential whiles the platform is in use. With these features, early recognition should prevent the occurrence of these harmful executions such as pivoting. This prediction can be based on a similar HMM approach (or LSTMs) in which angle, angular speed and angular acceleration forecasts harmful movement by differentiating on likelihood of sets of values that will result in unacceptable movement. Over a time-interval this can be interpreted as, if these values occur, what is the likelihood that the pose value (angle) exceeds the thresholded value in the near future. Without data to do so, an alternative could be the projection of acceleration into the near future to see if the threshold could be crossed with certain confidence, when the acceleration does not decrease sufficient.

New ways of exploring the realm of variations (good, compensatory, bad coordination) could be further developed in the virtual. Where VA's could produce for every degree of freedom (in this case an HMM), with a fixed duration, patterns of possibilities (by scanning the latent plane and sampling points of it). Then this multidimensional space could be explored by projections on avatars and sequentially demarcate clusters of clear error detection (therapist search). To start this, more about the flexibility per degree of freedom should be known and recordings can already aid in localizing some of the main clusters. Smart heuristics should aid in therapists exploring this flexible space.

Whit the creation of the assessment blueprint starts a new era. That of designing the appropriate levels of feedback and testing the trackability of progress during the recovery process of a patient. The blueprint provides an extensive load of measurements that still needs to be interpreted by the therapist and translated (automatically) into valuable feedback. Suggestions can be made in a similar way as in experiment II where sets are classified into the main types of errors. The proposed progress and plan module and the interactional model should therefore be further investigated to enable the cultivation of such expected impact. The challenges lie in determining in a dynamic fashion (considered the cognitive state of the user) how often, how much and on which aspect the feedback sound be provided. In addition, the construction of a database that contains recordings of therapeutic exercises (Mocap App) can be shared across multiple users whom are experimenting with the assessment. This collaborative framework will evoke communication between different therapists and speed up development in assessment techniques.

11 Conclusion

In this work creating an assessment method to automize the detection of quality in therapeutic exercises have been the objective. Various algorithms that can be used in time-series classification problems are explored and compared. With the wide range of developing platforms, commercially and in the scientific community, developing a proper interaction where users are being engaged seems to be the focus. In some, this interaction even extends to the possibility for therapists to create their own exercises. In most cases the assessment seems to be of a later focus and in multiple examples is kept primitive (binary classification e.g.). The yet vastly uncharted area of dynamic quality determination, derived from therapeutic tacit knowledge, is to a great degree explored in this thesis and has led to a novel HMM assessment method that could in practice be applied for any therapeutic exercise.

Representing human motion in a skeletonized fashion has shown to improve classification results and leads to and restricted descriptor that can express motion in therapist terminology (flexion, extension e.g.). For this to be possible the skeletonized images captured with a 3D motion capture devise is further refined so that anthropometrics do not interfere with the motion expression. Also as the final experiment demonstrated errors occur in the Skeletonization if subjects are not properly dressed and thus wearing tight clothing is an additional requirement whiles interacting. The results are a representation of motion in angular positioning of limbs and relative direction of movement of the subject.

Therapists input has created insight in which aspects are imported whiles assessing an exercise and the influence of additional factors in the thought process. The basic distinctions that are extracted from their reasoning while assessing an exercise are inspecting the range of motion, presence of compensatory elements (motion that should not be there), coordination of the movement (are body parts that move in the exercise synchronized) and finally the force expressed in the exercise (result of ROM, duration of max ROM state and anthropometrics). The first three metrics are incorporated into the research where for the later one a separate development of a mechanical/physical model would be suggested.

The HMM approach achieved real-time classification over various sizes of windows. This shows that error localization in an exercise can be used with this technique. As the HMMs are trained on data of other subjects to classify its execution and show overall excellent results, concluded is that with HMMs generalization is possible. Recording the bulk of different exercises in multiple sessions per experiment provided perspective on the impact when implementing the automated assessment. During these sessions, the need for intuitive applications to record and quickly create variations of HMMs became visible. This brought insight in the need for development of such tools when the platform will be finalized, and therapists need to work autonomously on exercise recording, model training and manual labelling (Prospective modules, chapter 9).

The later experiments show that training multiple HMMs on body segments creates more flexibility and potentially captures coordination as state transition paths can be compared as asynchronicity can be expressed as measure between them. The need for state duration and expectational transition paths needs to be integrated to further shape this coordination (in addition it will detect different types of compensation). As argued in the final experiment, not all compensations are a result of exceeding the speed distributions within states but additionally can have altered state transitions. Furthermore, the trained HMMs on healthy subjects creates similar classifications to the therapists labelling in classifying patient data. This concludes that it is possible that HMMs afford to generalize on a broad spectrum of subjects.

Finally, the proposed method is an extensive blueprint of meaningful measures (chapter 8, assessment methodology). These measures are created using a pipeline where in each step subtler errors can be detected. Starting with data collection of correct executions that need to be examined as such by a specialist is followed by a HMM learning process. The method can be used to provide targeted feedback as more evident errors should be tackled primarily. If the quality tracking of exercises does not contribute to an improvement in recovery, the developed technique can be utilized for analyses in fields where technique is an essential game changer such as motor skills applicable in sports, dancing and musical performance.

References

[1] B. Galna, G. Barry, D. Jackson, D. Mhiripiri, P. Olivier, and L. Rochester, "Accuracy of the Microsoft Kinect sensor for measuring movement in people with Parkinson's disease," *Gait & Posture*, vol. 39, no. 4, pp. 1062–1068, Apr. 2014.

[2] H. Jost, "Kinect-Based Approach to Upper Limb Rehabilitation," in *Modern Stroke Rehabilitation through e-Health-based Entertainment*, E. Vogiatzaki and A. Krukowski, Eds. Cham: Springer International Publishing, 2016, pp. 169–193.

[3] Q. V. Le, "Building high-level features using large scale unsupervised learning," in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, 2013, pp. 8595–8598.

[4] S. R. Kesler, N. J. Lacayo, and B. Jo, "A pilot study of an online cognitive rehabilitation program for executive function skills in children with cancer-related brain injury," *Brain Injury*, vol. 25, no. 1, pp. 101–112, Jan. 2011.

[5] D. Harley, G. Fitzpatrick, L. Axelrod, G. White, and G. McAllister, "Making the Wii at Home: Game Play by Older People in Sheltered Housing," in *HCI in Work and Learning, Life and Leisure*, vol. 6389, G. Leitner, M. Hitz, and A. Holzinger, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 156–176.

[6] A. Kübler, S. Kleih, and D. Mattia, "Brain Computer Interfaces for Cognitive Rehabilitation After Stroke," in *Converging Clinical and Engineering Research on Neurorehabilitation II*, vol. 15, J. Ibáñez, J. González-Vargas, J. M. Azorín, M. Akay, and J. L. Pons, Eds. Cham: Springer International Publishing, 2017, pp. 847–852.

[7] C. Barelle *et al.*, "KINOPTIM: A Tele-rehabilitation gaming Platform for Fall Prevention in the Elderly Community," *Int. J. of Health Research and Innovation*, vol. 2, no. 1, pp. 37–49, 2014.

[8] A. A. A. Timmermans *et al.*, "Sensor-Based Arm Skill Training in Chronic Stroke Patients: Results on Treatment Outcome, Patient Motivation, and System Usability," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 18, no. 3, pp. 284–292, Jun. 2010.

[9] A. Jacobs, A. Timmermans, M. Michielsen, M. Vander Plaetse, and P. Markopoulos, "CONTRAST: gamification of arm-hand training for stroke survivors," in *CHI'13 Extended Abstracts on Human Factors in Computing Systems*, 2013, pp. 415–420.

[10] V. F. S. Fook *et al.*, "Innovative platform for tele-physiotherapy," in *e-health Networking, Applications and Services, 2008. HealthCom 2008. 10th International Conference on*, 2008, pp. 59–65.

[11] R. Yeager, "An Automated Physiotherapy Exercise Generator," Citeseer, 2013.

[12] D. González-Ortega, F. J. Díaz-Pernas, M. Martínez-Zarzuela, and M. Antón-Rodríguez, "A Kinect-based system for cognitive rehabilitation exercises monitoring," *Computer Methods and Programs in Biomedicine*, vol. 113, no. 2, pp. 620–631, Feb. 2014.

[13] J. D. Westwood and others, "Real-time 3D avatars for tele-rehabilitation in virtual reality," *Medicine Meets Virtual Reality 18: NextMed*, vol. 163, p. 290, 2011.

[14] A. Tsavourelou *et al.*, "Telerehabilitation Solution Conceptual Paper for Community-Based Rehabilitation of Patients Discharged after Critical Illness," *International Journal of Telerehabilitation*, vol. 8, no. 2, pp. 61–70, 2016.

[15] S. Shamsuddin, "Development of human-robot interaction (HRI) methodology for autism rehabilitation using humanoid robot with a telerehabilitation platform," Universiti Teknologi MARA, 2015.

[16] M. Cortese, M. Cempini, P. R. de Almeida Ribeiro, S. R. Soekadar, M. C. Carrozza, and N. Vitiello, "A Mechatronic System for Robot-Mediated Hand Telerehabilitation," *IEEE/ASME Transactions on Mechatronics*, vol. 20, no. 4, pp. 1753–1764, Aug. 2015.

[17] E. Jovanov, A. Milenkovic, C. Otto, and P. C. De Groen, "A wireless body area network of intelligent motion sensors for computer assisted physical rehabilitation," *Journal of NeuroEngineering and rehabilitation*, vol. 2, no. 1, p. 1, 2005.

[18] S. Ponte, S. Gabrielli, J. Jonsdottir, M. Morando, and S. Dellepiane, "Monitoring game-based motor rehabilitation of patients at home for better plans of care and quality of life," 2015, pp. 3941–3944.

[19] M. Csikszentmihalyi, *Flow: the psychology of optimal experience*. New York: HarperPerennial, 1991.

[20] M. Callejas-Cuervo, R. M. Gutierrez, and A. I. Hernandez, "Joint amplitude MEMS based measurement platform for low cost and high accessibility telerehabilitation: Elbow case study," *Journal of Bodywork and Movement Therapies*, Sep. 2016.

[21] D. Pani *et al.*, "Home tele-rehabilitation for rheumatic patients: impact and satisfaction of care analysis," *Journal of Telemedicine and Telecare*, vol. 23, no. 2, pp. 292–300, Feb. 2017.

[22] B. Parmanto *et al.*, "VISYTER: Versatile and Integrated System for Telerehabilitation," *Telemedicine and e-Health*, vol. 16, no. 9, pp. 939–944, Nov. 2010.

[23] D. Antón, A. Goñi, and A. Illarramendi, "Exercise Recognition for Kinect-based Telerehabilitation:," *Methods of Information in Medicine*, vol. 54, no. 2, pp. 145–155, Oct. 2014.

[24] M. Caporuscio, D. Weyns, J. Andersson, C. Axelsson, and G. Petersson, "IoT-enabled Physical Telerehabilitation Platform," in *Proceedings of the International Workshop on Engineering IoT Systems: Architectures, Services, Applications, and Platforms*, 2017.

[25] R. N. Madeira, L. Costa, and O. Postolache, "PhysioMate-Pervasive physical rehabilitation based on NUI and gamification," in *Electrical and Power Engineering (EPE), 2014 International Conference and Exposition on*, 2014, pp. 612–616.

[26] J. Choi *et al.*, "Delivering an In-Home Exercise Program via Telerehabilitation: A Pilot Study of Lung Transplant Go (LTGO)," *International Journal of Telerehabilitation*, vol. 8, no. 2, pp. 15–26, 2016.

[27] A. Rothgangel, S. Braun, R. Smeets, and A. Beurskens, “Design and Development of a Telerehabilitation Platform for Patients with Phantom Limb Pain: A User-Centered Approach,” *JMIR Rehabilitation and Assistive Technologies*, vol. 4, no. 1, p. e2, Feb. 2017.
[28] B. Holm, K. Thorborg, H. Husted, H. Kehlet, and T. Bandholm, “Surgery-Induced Changes and Early Recovery of Hip-Muscle Strength, Leg-Press Power, and Functional Performance after Fast-Track Total Hip Arthroplasty: A Prospective Cohort Study,” *PLoS ONE*, vol. 8, no. 4, p. e62109, Apr. 2013.
[29] “Recovering from Hip Replacement Surgery, https://www.ucsfhealth.org/education/recovering_from_hip_replacement_surgery/,” 13-Jun-2017.
[30] D. Mangusan, “Physiotherapy Evaluation and Examination,” 08-May-2017. .
[31] J. J. A. van Boxtel, “Free Mocap Databases,” *3D (motion capture) databases, http://www.jeroenvanboxtel.com/MocapDatabases.html*.
[32] J. Shotton *et al.*, “Real-time human pose recognition in parts from single depth images,” *Communications of the ACM*, vol. 56, no. 1, p. 116, Jan. 2013.
[33] R. B. Rusu, G. Bradski, R. Thibaux, and J. Hsu, “Fast 3D recognition and pose using the viewpoint feature histogram,” in *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*, 2010, pp. 2155–2162.
[34] B. Peng, L. Zhang, and D. Zhang, “A survey of graph theoretical approaches to image segmentation,” *Pattern Recognition*, vol. 46, no. 3, pp. 1020–1038, Mar. 2013.
[35] Y. Yang and D. Ramanan, “Articulated Human Detection with Flexible Mixtures of Parts,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 12, pp. 2878–2890, Dec. 2013.
[36] M. Andriluka, S. Roth, and B. Schiele, “Pictorial structures revisited: People detection and articulated pose estimation,” in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, 2009, pp. 1014–1021.
[37] Y. Tian, R. Sukthankar, and M. Shah, “Spatiotemporal Deformable Part Models for Action Detection,” 2013, pp. 2642–2649.
[38] S.-C. Cheng and C.-K. Yang, “A fast and novel technique for color quantization using reduction of color space dimensionality,” *Pattern Recognition Letters*, vol. 22, no. 8, pp. 845–856, Jun. 2001.
[39] D. Nova and P. A. Estévez, “A review of learning vector quantization classifiers,” *Neural Computing and Applications*, vol. 25, no. 3–4, pp. 511–524, 2014.
[40] “Sparse Coding, http://ufldl.stanford.edu/wiki/index.php/Sparse_Coding,” 18-May-2017. .
[41] T. N. Sainath, B. Kingsbury, A. Mohamed, and B. Ramabhadran, “Learning filter banks within a deep neural network framework,” in *Automatic Speech Recognition and Understanding (ASRU), 2013 IEEE Workshop on*, 2013, pp. 297–302.
[42] B. Jansen, F. Temmermans, and R. Deklerck, “3D human pose recognition for home monitoring of elderly,” in *Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE*, 2007, pp. 4049–4051.
[43] A. F. Bobick and J. W. Davis, “The recognition of human movement using temporal templates,” *IEEE Transactions on pattern analysis and machine intelligence*, vol. 23, no. 3, pp. 257–267, 2001.
[44] F. Bogo, A. Kanazawa, C. Lassner, P. Gehler, J. Romero, and M. J. Black, “Keep it SMPL: Automatic estimation of 3D human pose and shape from a single image,” in *European Conference on Computer Vision*, 2016, pp. 561–578.
[45] L. Chen, H. Wei, and J. Ferryman, “A survey of human motion analysis using depth imagery,” *Pattern Recognition Letters*, vol. 34, no. 15, pp. 1995–2006, Nov. 2013.
[46] A. Yao, J. Gall, G. Fanelli, and L. V. Gool, “Does Human Action Recognition Benefit from Pose Estimation?,” 2011, p. 67.1-67.11.
[47] C. Barron and I. A. Kakadiaris, “Estimating anthropometry and pose from a single image,” in *Computer Vision and Pattern Recognition, 2000. Proceedings. IEEE Conference on*, 2000, vol. 1, pp. 669–676.
[48] K. Bennell *et al.*, “Hip and ankle range of motion and hip muscle strength in young female ballet dancers and controls.,” *British journal of sports medicine*, vol. 33, no. 5, pp. 340–346, 1999.
[49] L. Stathokostas, M. W. McDonald, R. M. D. Little, and D. H. Paterson, “Flexibility of Older Adults Aged 55?86 Years and the Influence of Physical Activity,” *Journal of Aging Research*, vol. 2013, pp. 1–8, 2013.
[50] X. K. Wei and J. Chai, “Modeling 3D human poses from uncalibrated monocular images,” in *Computer Vision, 2009 IEEE 12th International Conference on*, 2009, pp. 1873–1880.
[51] C. J. Taylor, “Reconstruction of articulated objects from point correspondences in a single uncalibrated image,” 2000, vol. 1, pp. 677–684.
[52] C. Wang, Y. Wang, Z. Lin, A. L. Yuille, and W. Gao, “Robust estimation of 3D human poses from a single image,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 2361–2368.
[53] A. Hernández-Vela *et al.*, “Probability-based Dynamic Time Warping and Bag-of-Visual-and-Depth-Words for Human Gesture Recognition in RGB-D,” *Pattern Recognition Letters*, vol. 50, pp. 112–121, Dec. 2014.
[54] I. Laptev, “On Space-Time Interest Points,” *International Journal of Computer Vision*, vol. 64, no. 2–3, pp. 107–123, Sep. 2005.
[55] I. Laptev, M. Marszalek, C. Schmid, and B. Rozenfeld, “Learning realistic human actions from movies,” 2008, pp. 1–8.
[56] G. V. Kale and V. H. Patil, “A Study of Vision based Human Motion Recognition and Analysis:,” *International Journal of Ambient Computing and Intelligence*, vol. 7, no. 2, pp. 75–92, Jun. 2016.
[57] J. K. Aggarwal and M. S. Ryoo, “Human activity analysis: A review,” *ACM Computing Surveys*, vol. 43, no. 3, pp. 1–43, Apr. 2011.
[58] S. Celebi, A. S. Aydin, T. T. Temiz, and T. Arici, “Gesture Recognition using Skeleton Data with Weighted Dynamic Time Warping.,” in *VISAPP (1)*, 2013, pp. 620–625.
[59] M. Müller, “Dynamic time warping,” *Information retrieval for music and motion*, pp. 69–84, 2007.
[60] J. Yamato, J. Ohya, and K. Ishii, “Recognizing human action in time-sequential images using hidden Markov model,” presented at the Proc. of IEEE Conference on Computer Vision and Pattern Recognition, 1992, pp. 379–385.
[61] Y. Du, W. Wang, and L. Wang, “Hierarchical recurrent neural network for skeleton based action recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1110–1118.

[62] G. Lefebvre, S. Berlemont, F. Mamalet, and C. Garcia, “BLSTM-RNN based 3D gesture classification,” in *International Conference on Artificial Neural Networks*, 2013, pp. 381–388.
[63] C. Finn, P. Abbeel, and S. Levine, “Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks,” *arXiv preprint arXiv:1703.03400*, 2017.
[64] V. Pavlovic and J. M. Rehg, “Impact of dynamic model learning on classification of human motion,” in *Computer Vision and Pattern Recognition, 2000. Proceedings. IEEE Conference on*, 2000, vol. 1, pp. 788–795.
[65] F. Zhou, F. De La Torre, and J. K. Hodgins, “1 Hierarchical Aligned Cluster Analysis (HACA) for Temporal Segmentation of Human Motion,” 2008.
[66] L. Xia, C.-C. Chen, and J. K. Aggarwal, “View invariant human action recognition using histograms of 3D joints,” in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2012 IEEE Computer Society Conference on*, 2012, pp. 20–27.
[67] X. Yang and Y. Tian, “Effective 3D action recognition using EigenJoints,” *Journal of Visual Communication and Image Representation*, vol. 25, no. 1, pp. 2–11, Jan. 2014.
[68] M. Capecci *et al.*, “Physical rehabilitation exercises assessment based on Hidden Semi-Markov Model by Kinect v2,” 2016, pp. 256–259.
[69] Feng Zhou and F. De la Torre, “Generalized Canonical Time Warping,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 2, pp. 279–294, Feb. 2016.
[70] M. Buhrmester, T. Kwang, and S. D. Gosling, “Amazon’s Mechanical Turk: A New Source of Inexpensive, Yet High-Quality, Data?,” *Perspectives on Psychological Science*, vol. 6, no. 1, pp. 3–5, Jan. 2011.
[71] Y. Rybarczyk *et al.*, “ePHoRt project: a web-based platform for home motor rehabilitation.,” *5th World Conference on Information Systems and Technologies. Madeira, Portugal.*, 2017.
[72] Y. Rybarczyk *et al.*, “Recognition of physiotherapeutic exercises through DTW and low-cost vision-based motion capture.,” *8th International Conference on Applied Human Factors and Ergonomics. Los Angeles, USA.*, 2017.
[73] IFAWC, O. Herzog, and H. Kenn, Eds., *Proceedings / IFAWC - 3rd International Forum on Applied Wearable Computing 2006: March 15 - 16, 2006 in Bremen*. Berlin Offenbach: VDE-Verl, 2006.
[74] O. Banos, M. Damas, H. Pomares, A. Prieto, and I. Rojas, “Daily living activity recognition based on statistical feature quality group selection,” *Expert Systems with Applications*, vol. 39, no. 9, pp. 8013–8021, Jul. 2012.
[75] U. Maurer, A. Smailagic, D. P. Siewiorek, and M. Deisher, “Activity Recognition and Monitoring Using Multiple Sensors on Different Body Positions,” 2006, pp. 113–116.
[76] J. Parkka, M. Ermes, P. Korpipaa, J. Mantyjarvi, J. Peltola, and I. Korhonen, “Activity Classification Using Realistic Data From Wearable Sensors,” *IEEE Transactions on Information Technology in Biomedicine*, vol. 10, no. 1, pp. 119–128, Jan. 2006.
[77] P. Besson, J. Richiardi, C. Bourdin, L. Bringoux, D. R. Mestre, and J.-L. Vercher, “Bayesian networks and information theory for audio-visual perception modeling,” *Biological Cybernetics*, vol. 103, no. 3, pp. 213–226, Sep. 2010.
[78] P. Besnard and Conference on Uncertainty in Artificial Intelligence, Eds., *Uncertainty in artificial intelligence: proceedings of the eleventh conference (1995), August 18 - 20, 1995, McGill University, Montréal, Québec, Canada*. San Francisco, Calif: Kaufmann, 1995.
[79] R. Kohavi and G. H. John, “Wrappers for feature subset selection,” *Artificial Intelligence*, vol. 97, no. 1–2, pp. 273–324, Dec. 1997.
[80] A. D. Calin, “Gesture Recognition on Kinect Time Series Data Using Dynamic Time Warping and Hidden Markov Models,” in *Symbolic and Numeric Algorithms for Scientific Computing (SYNASC), 2016 18th International Symposium on*, 2016, pp. 264–271.
[81] S. Riccadonna, G. Jurman, R. Visintainer, M. Filosi, and C. Furlanello, “DTW-MIC Coexpression Networks from Time-Course Data,” *PLOS ONE*, vol. 11, no. 3, p. e0152648, Mar. 2016.
[82] J. F.-S. Lin and D. Kulic, “Online Segmentation of Human Motion for Automated Rehabilitation Exercise Analysis,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 22, no. 1, pp. 168–180, Jan. 2014.
[83] J. Fiosina and M. Fiosins, “Resampling based modelling of individual routing preferences in a distributed traffic network.,” *International Journal of Artificial Intelligence 12, 79-103.*, 2014.
[84] Smyth, P., 1996, "Clustering using Monte Carlo cross-validation". Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining, 2-4 Aug. 1996, Portland, Oregon, USA.
[85] “Vrkalovic, S., Teban, T.A., Borlea, I.D., 2017, Stable Takagi-Sugeno fuzzy control designed by optimization, International Journal of Artificial Intelligence 15, 17-29.”
[86] R. D. Baruah and P. Angelov, “DEC: Dynamically Evolving Clustering and Its Application to Structure Identification of Evolving Fuzzy Models,” *IEEE Transactions on Cybernetics*, vol. 44, no. 9, pp. 1619–1631, Sep. 2014.
[87] H. Zheng, R. Wang, Y. Wang, and W. Zhu, “Fault diagnosis of photovoltaic inverters using hidden Markov model,” 2017, pp. 7290–7295.
[88] L. Xue, J. Yin, Z. Ji, and L. Jiang, “A Particle Swarm Optimization for Hidden Markov Model Training,” 2006.
[89] A. Mahapatra, T. K. Mishra, P. K. Sa, and B. Majhi, “Human recognition system for outdoor videos using Hidden Markov model,” *AEU - International Journal of Electronics and Communications*, vol. 68, no. 3, pp. 227–236, Mar. 2014.
[90] M. Baratchi, N. Meratnia, P. J. M. Havinga, A. K. Skidmore, and B. A. K. G. Toxopeus, “A hierarchical hidden semi-Markov model for modeling mobility data,” 2014, pp. 401–412.
[91] Q. Wang, Y. Xu, Y.-L. Chen, Y. Wang, and X. Wu, “Dynamic hand gesture early recognition based on Hidden Semi-Markov Models,” 2014, pp. 654–658.
[92] J. Yang, M. N. Nguyen, P. P. San, X. Li, and S. Krishnaswamy, “Deep Convolutional Neural Networks on Multichannel Time Series for Human Activity Recognition.,” in *IJCAI*, 2015, pp. 3995–4001.
[93] V. Pterneas, “Kinect Joint Rotation,” *Kinect Joint Rotation – The Definitive Guide, https://www.codeproject.com/Articles/1189463/Kinect-Joint-Rotation-The-Definitive-Guide*, 2017.

Appendix

Appendix A – Recording and Training application

In this appendix, the developed applications that have been used in the experiments are shorty discussed. The figure description will be the guidance in this explanation.

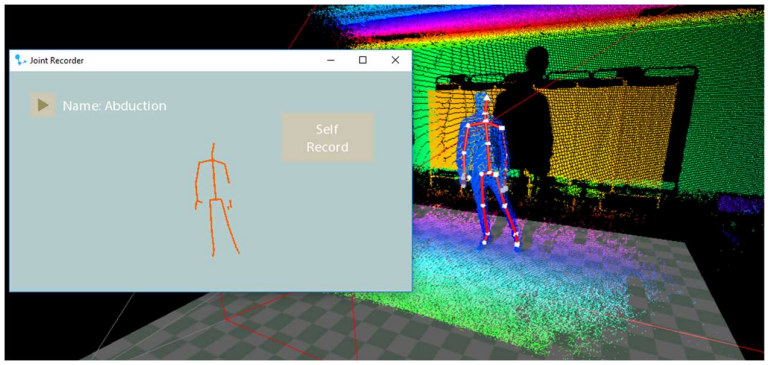


Figure A.1 Visualization of the recording application and the 3D environment with the detected skeleton (only the first skeleton in the scene will be recorded in the application). The name of the exercise can be entered and in the recording repository the unique recording name is the entered name plus an accumulated number (one higher than the highest number found).

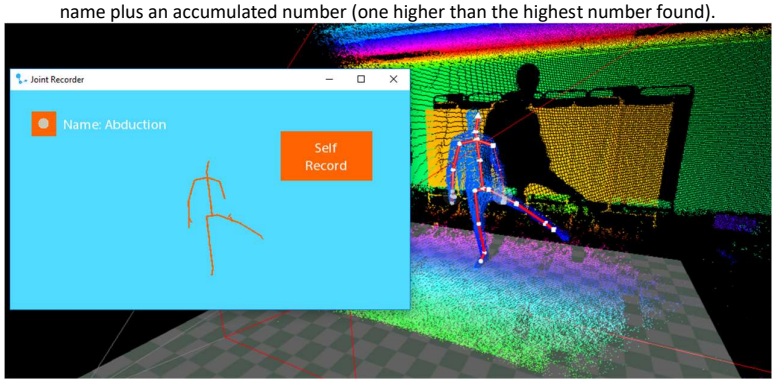


Figure A.2 When the play button is pressed, the screen changes colour to visually communicate to the person executing the recording the status. If there are any other skeletons in the image, during the recoding they will not be visualized nor recorded. When the button is pressed a sound is played to also indicate the status to the subject that is performing the exercise.

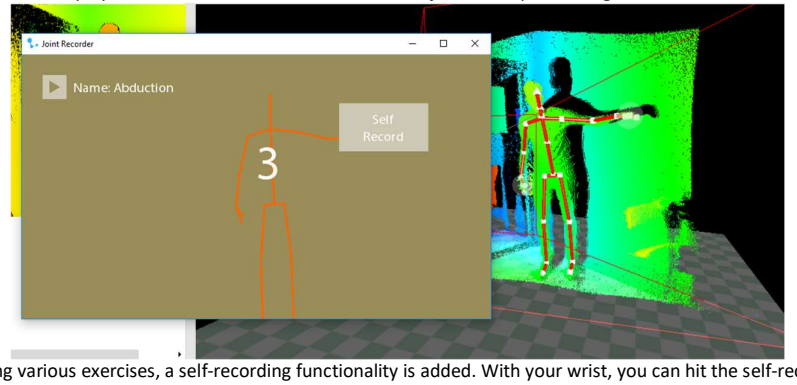


Figure A.3 Whiles testing various exercises, a self-recording functionality is added. With your wrist, you can hit the self-record button when after this a could down appears. After this count down the recording starts. When finishing the recording the last second of data is removed as this only corresponds to raising the arm to the button again.

An application was created to enable the therapists to easily create new assessment models that could be applied on any kind of rehabilitation exercises. Csv files can be loaded by clicking an HMM button. Before doing so the features that are going to be used in training need to be assigned by typing them in the features section (Figure A.4). To analyse the optimal amount of states that a dataset requires, Bayesian Information Criteria (BIC) can be performed. This creates a score using cross validation for states 2-10 (Figure A.5, as a hidden Markov model exists at least out of 2 states). The amount of states to be used during the training can be declared in the states section. When all the parameters are set, the training button appears, and the training can take place. When the training is done, the models can be saved (Figure A.6).

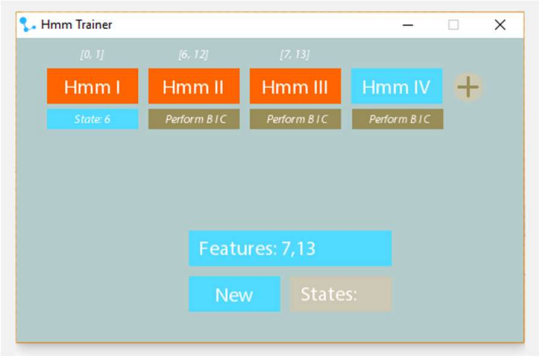


Figure A.4 Data can be loaded by clicking on a HMM button (it will turn orange when data is loaded), then the corresponding columns of the data that are going to be used in training can be selected in the features box. Clicking on the plus creates a possibility to train multiple HMMs in one session. The Model names are derived from the first part of the name of the loaded data files, e.g. HMMI-Abduction, HMMII-Abduction etc. for exercises starting with abduction.

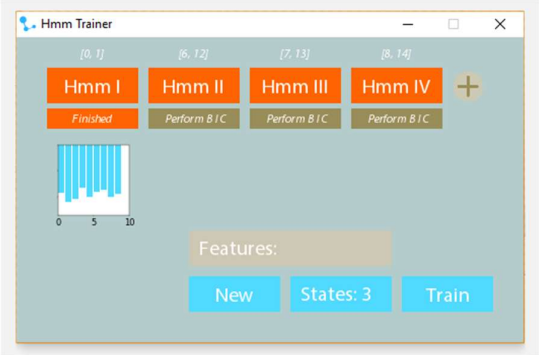


Figure A.5 When the button perform BIC is clicked the score will show up in a graph where lower scores correspond to an optimal amount of states. The scores are calculated using a 10-crossfold validation where the ultimate value shown in the graph is an average of these 10 validations.

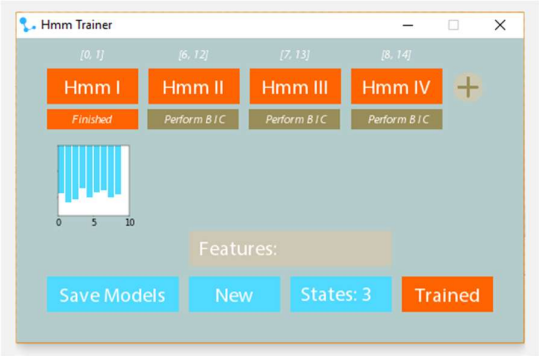


Figure A.6 When selecting the appropriate amount of states (per HMM) the train button can be clicked. When the training is completed, a new bottom will appear. This button lets the used save the models. The models will directly be saved into the directory of the used data. The model is a .pkl file that can be easily loaded within a python environment to perform classification.

Appendix B – Pseudo Code

```
Def Sample_from_models(models):
        S_numbers = 10000
        Sample_data = zeros(S_numbers,Features,States,size(models))
        For mod_num, model in enumerate(models):
                Predicted = [statesequence_O1, ..., statesequence_On]
                Percentage_Values = Histogram(Predicted,Amount_of_States)/size(Predicted)
                For n in States:
                        New_Samples = Model[n].sample(Percentage_Values[n]*S_numbers)
                        Sample_data[:,:,n,mod_num].append(New_Samples)
        Return Sample_data
```

Pseudo code 1. Finding the percentage of each states general contribution to create a sampled data distribution of a given HMM.

```
# Optimal Feature comparing two models
Def Optimal_Feature(models):
        sample_Hist = zeros(Features,30,models)
        For F in Features:
                sample_Hist(F,:,:) = Histogram(Sample_data(:,F,:,[models]),30)
        # keep only the bins with more than 1%
        sample_Histogram[sample_Histogram<S_numbers/100]= 1
        ratio_Matrix = Devide(sample_Hist(:,:,models[0]),sample_Hist(:,:,models[1]))
        Invert_index                    =                    where(ratio_Matrix>1)
        ratio_matrix[Invert_index] = 1/ratio_Matrix[Invert_index]
        ratio_matrix.sum(axis=1)
        Return Optimal_Feature = Argmin(ratio_Matrix)
```

Pseudo code 2. Finding the optimal feature, in which the class differentiation between 2 HMMs is the highest.

Appendix C – Therapist labelling

The therapist labelling of the 9 subjects and the 2 patients can be found here. Each subject executed 5 movements, in order: Hip abduction, Hip extension, step forward; sideways; backwards (SSB), Hip flexion sitting and Hip flexion. Each coloured column corresponds to these exercises respectively (Table C.1-Table C.7). Every execution is labelled on 4 different traits being: ROM, Coordination, Compensation and Force. These are respectively represented as individual rows per coloured block. For the SSB movement the ROM is not evaluated. The number of each execution keeps track of the labelling (excellent, good or bad) per trait per exercise. Every therapist's labelling is shown in separate columns (with the corresponding therapist name).

Table C.1 Therapist assesment of the 5 different exercises (8 executions per exercise) on subject 1.

Danilo			***Arian***			***Andrea***			***Yadira***			***Katherine***		
excellent	*good*	*bad*	*excellent*	*good*	*bad*	*excellent*	*good*	*bad*	*excellent*	*good*	*bad*	*excellent*	*good*	*bad*
12345678			12345678			12345678			12345678			12345678		
12345678			12345678			12345678			12345678			12345678		
12345678			12345678			12345678			12345678			12345678		
12345678			12345678			12345678			12345678			12345678		
12345678			12345678			12345678			12345678			12345678		
12345678			124678	35		12345678			1245678	3		12345678		
1245678	3		1245678	3		124567	38		1245678	3		124567	38	
12345678			12345678			12345678			12345678			12345678		
12345678			1234568	7		12345678			5678	1234		12345678		
12345678			12345678			12345678			12345678			1234678	5	
12345678			12345678			12345678			12345678			12345678		
12345678			12345678			12345678			12345678			12345678		
12345678			12345678			12345678			12345678			12345678		
234568	17		12345678			145678	23		1245678	3		234578	16	
12345678			12345678			12345678			12345678			12345678		
12345678			12345678			12345678			12345678			12345678		
12345678			245678	13		12345678			12345678			1245678	3	
12345678			12345678			12345678			1234578	6		12345678		
12345678			12345678			12345678			12345678			12345678		

Table C.2 Therapist assesment of the 5 different exercises (8 executions per exercise) on subject 2.

Danilo			***Arian***			***Andrea***			***Yadira***			***Katherine***		
excellent	*good*	*bad*	*excellent*	*good*	*bad*	*excellent*	*good*	*bad*	*excellent*	*good*	*bad*	*excellent*	*good*	*bad*
12345678			12345678			12345678			12345678			12345678		
1234568	7		158	23467		1234578	6		3468	1257		12345678		
134578	26		12345678	4		234568	17		34578	126		23468	157	
12345678			12345678			12345678			12345678			12345678		
12345678			12345678			12345678			12345678			12345678		
12345678			35	124678		123467	58		237	14568		123578	46	
1234567	8		1235678	4		123578	46		2367	1458		235678	14	
12345678			12345678			12345678			12345678			12345678		
12345678			14578	236		1235678	4		678	12345		1234678	5	
12345678			12345678			12345678			678	12345		12345678		
12345678			12345678			12345678			12345678			12345678		
12345678			35678	124		12345678			345678	12		12345678		
12345678			12345678	2		12345678			345678	12		123678	45	
12345678			12345678			1235678	4		345678	12		134678	25	
12345678			12345678			12345678			345678	12		12345678		
12345678			12345678			12345678			12345678			12345678		
12345678			134567	28		12345678			12345678			12345678		
12345678			1345678	2		12345678			12345678			12345678		
12345678			12345678			12345678			12345678			12345678		

Table C.3 Therapist assesment of the 5 different exercises (8 executions per exercise) on subject 3.

Danilo			Arian			Andrea			Yadira			Katherine		
excellent	*good*	*bad*	*excellent*	*good*	*bad*	*excellent*	*good*	*bad*	*excellent*	*good*	*bad*	*excellent*	*good*	*bad*
12345678			12345678			12345678			2345678	1		12345678		
12345678			134578	26		12345678			2468	1357		1245678	3	
1345678	2		1345678	2		2678	1345		2468	1357		23478	156	
12345678			12345678			12345678			12345678			12345678		
12345678			12345678			12345678			1245678	3		12345678		
12345678			1234678	5		12345678			1234568	7		12345678		
1345678	2		12345678			12345678			12345678			12345678		
12345678			12345678			12345678			12345678			12345678		
12345678			145678	23		12345678			12345678			12345678		
12345678			1234568	7		1245678	3		12345678			12345678		
12345678			12345678			12345678			12345678			12345678		
12345678			2345678	1		12345678			2345678	1		12345678		
12345678			234678	15		12345678			2345678	1		23568	147	
12345678			1234568	7		235678	14		234568	17		12345678		
12345678			12345678			12345678			2345678	1		12345678		
12345678			12345678			12345678			12345678			12345678		
12345678			24568	137		12345678			245678	13		23568	147	
12345678			12345678			345678	12		24568	137		24568	137	
12345678			12345678			12345678			12345678			12345678		

Table C.4 Therapist assesment of the 5 different exercises (8 executions per exercise) on subjects 4 (left) and 5 (right).

Danilo			Andrea			Katherine		
excellent	*good*	*bad*	*excellent*	*good*	*bad*	*excellent*	*good*	*bad*
12345678			12345678			12345678		
12345678			1345678	2		45678	123	
1235678	4		1345678	2		34678	125	
12345678			12345678			12345678		
12345678			12345678			12345678		
12345678			12345678			1234568	7	
12345678			12345678			12345678		
12345678			12345678			12345678		
12345678			12345678			12345678		
12345678			12345678			12345678		
12345678			12345678			12345678		
12345678			12345678			12345678		
12345678			12345678			12345678		
12345678			12345678			12345678		
12345678			12345678			12345678		
12345678			12345678			12345678		

Danilo			Arian		
excellent	*good*	*bad*	*excellent*	*good*	*bad*
12345678			12345678		
12345678			34578	126	
12345678			12345678		
12345678			12345678		
12345678			12345678		
12345678			1345678	2678	
12345678			12345678		
12345678			12345678		
12345678			12345678		
12345678			12345678		
12345678			12345678		
12345678			12345678		
12345678			12345678		
1345678	2		2345678	1	
1345678	2		1345678	2	
12345678			12345678		

Table C.5 Therapist assesment of the 5 different exercises (8 executions per exercise) subjects 6 (left) and 7 (right).

Danilo			Arian		
excellent	*good*	*bad*	*excellent*	*good*	*bad*
12345678			245678	13	
12345678			123478	56	
12345678			12345678		
12345678			12345678		
12345678			12345678		
12345678			12345678	12	
12345678			12345678		
12345678			12345678		
1235678	4		578	12346	
1235678	4		1235678		4
12345678			12345678		
12345678			12345678		
12345678			1235678	4	
12345678			12345678		
12345678			12345678		
12345678			2345678	1	
12345678			12345678		
12345678			12345678		

Danilo			Arian		
excellent	*good*	*bad*	*excellent*	*good*	*bad*
12345678			123456	78	
12345678			134578	26	
12345678			12345678		
12345678			12345678		
12345678			12345678		
12345678			345678	12	
12345678			12345678		
12345678			12345678		
12345678			23456	178	
123456	78		1234568	7	
12345678			12345678		
12345678			12345678		
12345678			35678	124	
12345678			12345678		
12345678			12345678		
12345678			12345678		
12345678			12345678		
12345678			12345678		

Table C.6 Therapist assesment of the 5 different exercises (8 executions per exercise) on subjects 8 (left) and 9 (right).

Danilo			Arian		
excellent	*good*	*bad*	*excellent*	*good*	*bad*
12345678			12345678		
234567	18		234567	18	
1234567	8		1234567	8	
12345678			12345678		
12345678			12345678		
12345678			1234567	8	
12345678			12345678		
12345678			12345678		
12345678			234567	18	
12345678			12345678		
12345678			12345678		
12345678			12345678		
12345678			134578	26	
12345678			12345678		
12345678			12345678		
12345678			12345678		
12345678			2345678	1	
12345678			12345678		
12345678			12345678		

Danilo			Arian		
excellent	*good*	*bad*	*excellent*	*good*	*bad*
12345678			12345678		
12345678			678	12345	
12345678			12345678		
12345678			12345678		
12345678			12345678		
12345678			234568	17	
12345678			12345678		
12345678			12345678		
1234567	8		1234578	6	
1234567	8		1234567	8	
12345678			12345678		
12345678			12345678		
12345678			12345678		
12345678			12345678		
12345678			12345678		
12345678			12345678		
12345678			234678	15	
12345678			12345678		
12345678			12345678		

Table C.7 Therapist assesment of the 5 different exercises (8 executions per exercise) on patients 1 (left) and 2 (right).

Danilo			Galo		
excellent	*good*	*bad*	*excellent*	*good*	*bad*
12345678			78	123456	
1234568	7		12345678		
12345678	45		12345678	5	4
12345678			12345678		
12345678				3467	1258
12345678			3567	1258	
12345678			12345678		
12345678			12345678		
			2456	1	38
1235678	4		12345678		
1235678	4		12345678		
12345678			12345678		
5678	1234		6	1234578	
345678	12		12345678		
35678	124		3456	128	7
12345678			12345678		
134567	2		2345678	1	
1234567			12345678		
1234567			12345678	1	
1234567			12345678		

Danilo			Galo		
excellent	*good*	*bad*	*excellent*	*good*	*bad*
12345678			678	12345	
12345678			678	12345	
12345678			678	12345	
12345678			678	12345	
678	12345		678	12345	
678	12345				
678	12345				
678	12345				
				1	
			8	1234567	
			678	12345	

Appendix D – Recording subjects

Here an impression of a recording session is visualized. The subjects as seen by the camera and the corresponding skeleton representation are shown.

Figure D.1 A healthy subject performing four different exercises, from the left top: Hip Abduction, SSB, Hip Extension and Hip Flexion.

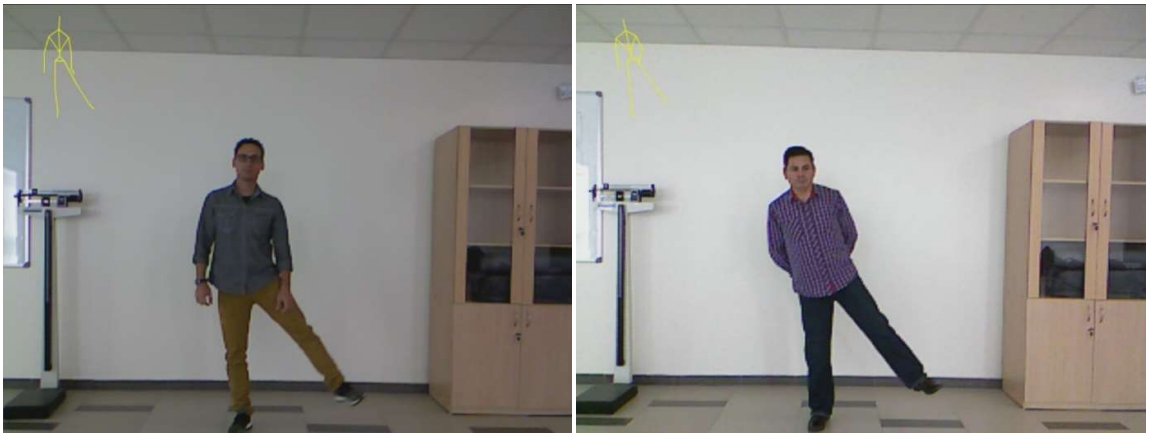


Figure D.2 2 heathy male Subjects performing Hip Abduction.

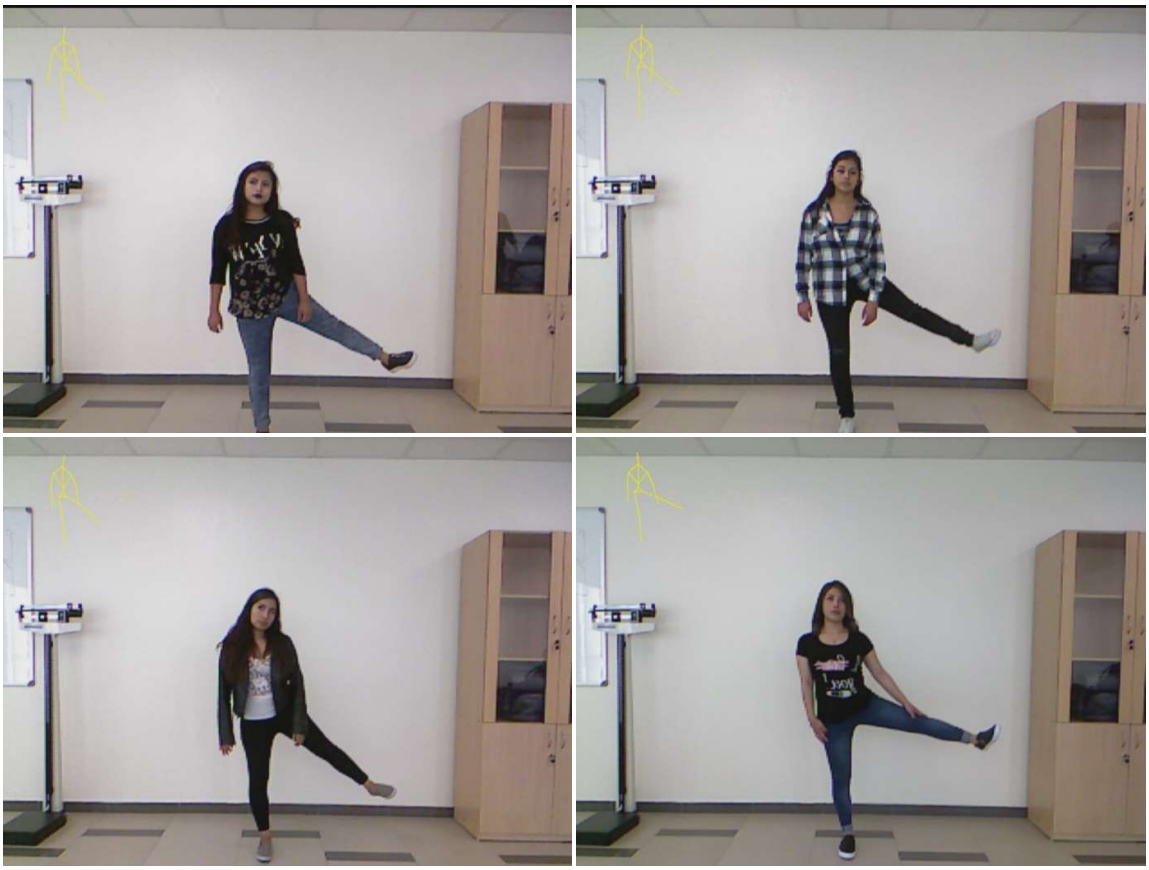


Figure D.3 4 healthy female Subjects performing Hip Abduction.

Appendix E – Wekinator & Browser development

A simple test was executed to see if of-the-shelf solutions could aid in the processes of automatic assessment. The Wekinator tool (http://www.wekinator.org/) applies DTW where first training data can be send to the application via OSC and later a classification can be fired from within the application on a live data stream. In Figure E.1 a screenshot of the Wekinator application and the recording application can be seen. Additional work has been done on connecting the Kinect to the server application of the project. The result is a data stream that utilizes the web sockets protocol to communicate to the server and a local Node.js application is installed to obtain the skeleton data. Then with JavaScript the data can be visualized (Figure E.2) in the browser so that eventually the end user does not need to install additional software when using the platform.

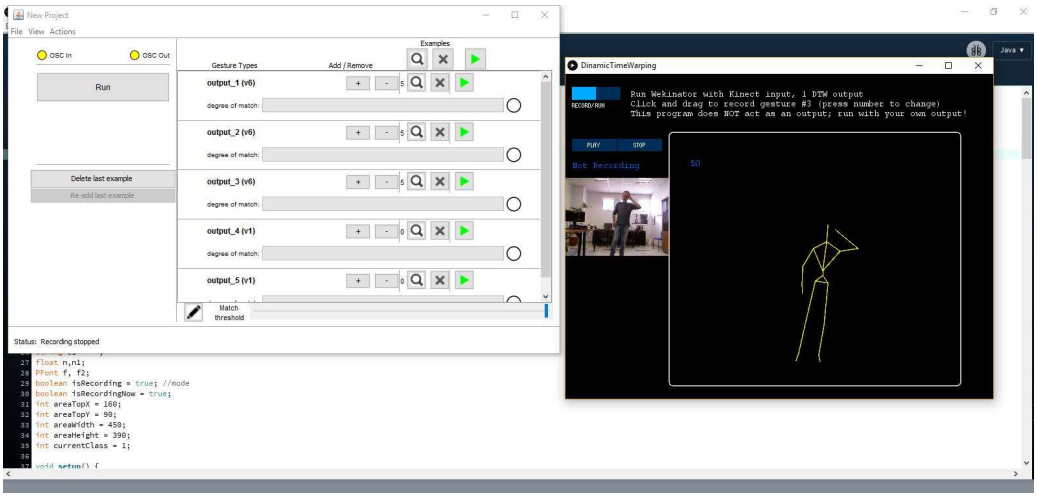


Figure E.1 Left the Wekinator application and on the right the recording application (Kinect + processing) that sends the data to Wekinator to train a DTW model or to classify the data.

Figure E.2 Kinect's data visualized in the browser using Node.js