

**COUPLING OF AN OBJECTIVE AND QUANTIFIABLE
METHODOLOGY FOR ASSESSING UPPER-BODY
MOVEMENTS WITH VIRTUAL REALITY GAMING
PLATFORMS**

A Thesis
Presented to
The Academic Faculty

by

Sergio García-Vergara

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Electrical and Computer Engineering

Georgia Institute of Technology
May 2017

Copyright © 2017 by Sergio García-Vergara

COUPLING OF AN OBJECTIVE AND QUANTIFIABLE METHODOLOGY FOR ASSESSING UPPER-BODY MOVEMENTS WITH VIRTUAL REALITY GAMING PLATFORMS

Approved by:

Professor Ayanna M. Howard, Advisor
School of Electrical and Computer
Engineering
Georgia Institute of Technology

Professor Aaron Ames
School of Mechanical Engineering
Georgia Institute of Technology

Professor Patricio Vela
School of Electrical and Computer
Engineering
Georgia Institute of Technology

Professor Yu-Ping Chen
School of Nursing and Health
Professions
Georgia State University

Professor Pamela Bhatti
School of Electrical and Computer
Engineering
Georgia Institute of Technology

Date Approved: January 2017

To my parents and sister, without whom none of this would be possible.

ACKNOWLEDGEMENTS

I greatly appreciate all the help and support I've received over the years. I would like to thank my advisor, Dr. Ayanna Howard, who has provided me with invaluable advise throughout my time at GaTech. It is a privilege to call myself her student. I would also like to thank my committee members, (Dr. Aaron Ames, Dr. Patricio Vela, Dr. Yu-Ping Chen, and Dr. Pamela Bhatti) for their invaluable time and attention. I definitely would not be here if it weren't for my parents, Raquel and Carmelo. Thank you for your unconditional love and support, and for the constant reminders that the finish line is just around the corner. To my GaTech colleagues and fellow HumAnS, thank you for your patience and friendship. Finally, I am grateful to my friends in Puerto Rico and Atlanta who have become an extension of my family.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

According to the Center for Disease Control, cerebral palsy is the most common motor disability in the US. Physical therapy helps alleviate the burden of the symptoms, reduce the rate of deterioration, and enhance functional capacity. Researchers have observed that, due to lack of motivation and reduced engagement, there is a non-trivial amount of patients that do not perform the recommended in-home therapy exercises. Lack of compliance limits greater improvement due to interrupted training. In this era of novel technological advances, virtual reality serious gaming systems have been designed to address this problem by creating fun and engaging scenarios that increase user motivation. However, existing systems do not fully mimic the interactions between patients and their trained therapists. Namely, they do not employ assessment of the user's kinematic performance to objectively keep track of their progress, nor do they provide the necessary corrective feedback.

External feedback of performance is an important component of therapy. This is especially true for children who have cerebral palsy since, in general, they lack the intrinsic definition of what a “good movement” is, which they need in order to compare their current performance and better determine an approach to correct it. As such, the purpose of this dissertation is to fill in the gap by designing, developing, and validating a more robust system that can provide the necessary targeted corrective feedback as a function of the objective assessment of the user's kinematic performance.

We began by developing our own virtual reality gaming platform that not only maintains user engagement during game play, but also allows for the individualization of the intervention protocol to tailor the interaction to the user's needs and

body dimensions. We conducted a user study to determine the level of users' interest/enjoyment during game play, and found that our system can maintain its users' interest and enjoyment while minimizing pressure and tension. Moving forward, we embedded in the system an objective and quantifiable methodology that keeps track of the users' performance relative to a set of kinematic parameters we defined to mathematically describe an individual's upper-body movements. We made sure to validate the overall system's efficacy and efficiency by quantifying the sensing methods, and examining the system's test-retest reliability as well as its feasibility of being used as an evaluation tool. We then developed a machine learning approach to identify an individual's kinematic level. This is the first step towards allowing for the capability of continuously providing corrective feedback as a function of the user's abilities. Results from our final user studies show that, via interactions with our system (i.e. receiving appropriate corrective feedback from a robotic playmate), users can 1) reach specific performance reference values for a given kinematic parameter, and 2) decrease their movement time after a training session with a robotic playmate. This results support our claim that a more robust system that better mimics the interactions between a patient and their trained therapists will increase the efficacy of the intervention protocol and the user's rate of improvement as compared to interactions with existing serious gaming systems.

The conclusions of this dissertation are based on the results of numerous user studies with a total of almost 250 participants, including able-bodied adults and teenagers, typically developing children, and children with cerebral palsy. This dissertation presents data to support the claim that our system has the potential to serve as part of various physiotherapy protocols for individuals who have some form of motor skills disorder towards improving the efficacy of the protocols and increasing their rate of improvement.

CHAPTER I

INTRODUCTION

1.1 Motivation

For individuals who have some form of motor skills disorder, physical therapy interventions are recommended as part of their rehabilitation protocol to increase mobility and function, and reduce further development of the symptoms. The benefits of physical therapy have been well documented [?, ?, ?, ?, ?, ?]. Unfortunately however, previous studies have found that there is a non-trivial amount of patients that do not comply with their in-home training exercises for various reasons. In general, family- and child-centered therapy care often incorporates some form of play [?, ?, ?], based on results showing that children are more relaxed, intrinsically motivated, and actively engaged, which are all behaviors conducive to learning [?]. Based on these findings, virtual reality (VR) serious gaming systems have been designed to increase user motivation and compliance by, ultimately, masking the tediousness of performing repetitive exercises with fun and engaging scenarios.

Recent studies have documented some benefits to using VR gaming systems for physiotherapy purposes. For example, they have been shown to be effective motivational tools [?], as well as functional tools for training arm function [?]. However, existing systems do not fully mimic the interactions between patients and their trained therapists. Namely, existing systems do not employ objective assessment of the users' kinematic performance nor do they provide the necessary corrective feedback users need to efficiently improve their performance.

This dissertation focuses on the design, development, and validation of a novel in-home VR gaming system for the upper-body rehabilitation of individuals who have

some form of motor skills disorder. The goal is to develop a robust system that not only increases user motivation and engagement, but also has the capability of continuously providing targeted corrective feedback as a function of the real-time objective and quantifiable assessment of the user’s kinematic performance. In the pursuit of this effort, we have designed and developed a system that: 1) can be used outside of the clinical setting, 2) allows for the individualization of the intervention protocol to meet the specific needs of the user, 3) increases user motivation to comply with their intervention protocols, 4) keeps track of the user’s kinematic performance via an objective and quantifiable assessment methodology, and 5) adapts interaction with the system by continuously providing targeted corrective feedback as a function of the user’s kinematic performance.

1.2 Objectives

1.2.1 Thesis Statement

An in-home virtual reality serious game coupled with an objective and quantifiable assessment methodology is a feasible and effective approach for inducing changes in users’ kinematic behavior in real-time.

1.2.2 Contributions

To begin this work, we first developed our own virtual reality (VR) gaming platform: the *Super Pop VRTM* game (described in Chapter ??). During the design process, we adhere to specific characteristic features of existing VR serious gaming systems: 1) individuals can interact with the system outside the clinical setting, for example, in the comfort of their homes, and 2) they allow for the individualization of the interactions. We address the first feature by using the KinectTM camera from Microsoft as the sensing mechanism because of its portability. We address the second feature by adding the capability for the patient’s therapist or caregiver to be able to select the

game settings that are most appropriate for the user’s intervention protocol via user-friendly graphic user interfaces. Moreover, we made sure that our system addresses the lack of compliance patients experience by disguising intervention protocols with an entertaining game. We also present the results of a user study conducted to validate the capability of our system to maintain the users’ interest and enjoyment while minimizing pressure and tension during game play. This led to our first contribution:

1. *Developed a virtual reality serious game for the in-home setting that allows for the individualization of the therapy protocols, and increases user engagement and compliance with their intervention protocols.*

Moving forward, we developed an objective and quantifiable assessment methodology used to keep track of the user’s kinematic performance. To construct the baseline models against which an individual’s performance is compared to, instead of going through the time-consuming and potentially tedious process of collecting human data, we applied a kinematic model that mimics the human arm. The model generates ground truth values for a set of kinematic parameters we defined to mathematically describe an individual’s upper-body movements. Thus, in Chapter ??, we also describe the user studies we conducted to validate the kinematic model as well as the baseline it generates, and to evaluate our system’s efficacy and efficiency. This research led to our second contribution:

2. *Defined a list of kinematic parameters that, coupled with an applied kinematic model of the human arm, mathematically describe an individual’s upper-body movements to generate baseline movement trajectories as a function of the current movement task and the user’s upper-body dimensions.*

As previously mentioned, our system uses the KinectTM camera as its sensing method to eliminate its limitation to the clinical setting due to the lack of portability of

current state-of-the-art motion-capture systems. As such, we found it important to quantify the Kinect’s performance. After confirming that our system can yield similar performance as highly accurate motion-capture systems, we examined the system’s feasibility of being used as an evaluation tool for assessing kinematic performance to validate our claim that our algorithm can yield results that correlate with those of standard clinical assessment methodologies. Moreover, when developing a new evaluation tool, it’s also important to examine its test-retest reliability within and between days. Thus, in Chapters ?? and ??, we describe the user studies we conducted to evaluate and examine our system’s efficiency and efficacy. This research led to our third contribution:

3. Validated the feasibility of using the Super Pop VRTM game as a reliable and accurate evaluation tool to measure individuals’ reaching kinematics by quantifying the accuracy of the sensing methods, examining the system’s feasibility of being used as an evaluation tool, and examining the test-retest reliability of the overall system.

External corrective feedback is most efficient when provided relative to the user’s abilities. As such, for our system to have the capability of providing targeted corrective feedback, it first needs the capability to identify the user’s current kinematic level. Continuous identification of the user’s kinematic level allows the system to autonomously select the most appropriate baseline model such that the feedback is targeted relative to his/her kinematic performance at any given point in time. Thus, in Chapter ??, we compare the performances of the different pattern recognition classification models we implemented, and discuss the approach we developed that best separates the classes of interest. This research led to our fourth contribution:

4. Developed an approach to classify an individual’s kinematic performance class as a function of their upper-body reaching kinematics.

In a previous study, we found that the most effective manner to provide corrective feedback - in terms of an individual performing closer to the feedback cues - is by providing a combination of verbal and nonverbal cues. Taking these results into consideration, we conducted two users studies to determine if, via interactions with our system, individuals can: 1) modify their kinematic behavior such that they can reach an individualized performance goal, and 2) improve their kinematic performance while continuously receiving corrective feedback and then maintaining said performance after the feedback is removed. This research, which is described in Chapter ??, led to our fifth and final contribution:

5. Conducted human-robot interaction studies to examine the system's ability to prompt users to adapt and modify their kinematic behavior as a function of the provided corrective feedback.

1.2.3 Complete System

The combination of all the contributions resulting from this dissertation yields our final system as described in the diagram shown in Figure ?. The diagram describes the functionality of the complete system. Each box in the diagram makes reference to the chapter in the dissertation that addresses the corresponding process.

1.3 Dissertation Outline

The dissertation is organized as follows:

- **Chapter ?** summarizes the work that has been done to improve the efficacy of physical therapy including general descriptions of 1) previously developed virtual reality (VR) systems and serious gaming platforms used for rehabilitation, and 2) currently used assessment methodologies for keeping track of individuals' progress throughout their intervention protocols.

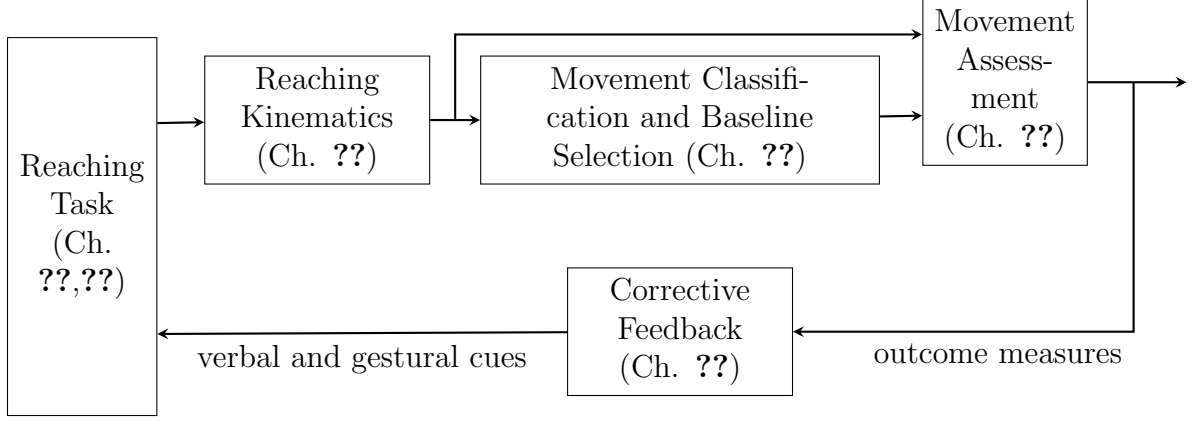


Figure 1: Diagram that describes the complete system. Each box makes reference to the chapter that addresses the corresponding process.

- **Chapter ??** describes in detail the work we completed on the development of the *Super Pop VRTM* system, a VR serious game to be included in the rehabilitation protocol for children who have cerebral palsy. The filtering methods used to improve tracking and the analysis of upper-body movements are addressed, and the algorithm used to select appropriate evaluation trajectories is introduced. Finally, the chapter presents the engagement study conducted to determine how engaging users perceive the game to be.
- **Chapter ??** presents the complete approach on using the Penalized Manipulator Jacobian to model the kinematics of the human arm as well as the definitions of the selected kinematic parameters to mathematically describe an individual's upper-body movements. The studies conducted to validate the kinematic model are also discussed.
- **Chapter ??** discusses the work done on the validation of the proposed system. We quantified the accuracy of the sensing method relative to a marker-based motion capture system, tested the *Super Pop VRTM* system with adults, typically developing kids, and children who have cerebral palsy, and validated the baseline generated by our kinematic model by comparing to a baseline generated

with human data.

- **Chapter ??** compares the different machine learning methodologies researched towards classifying upper-body movements as typical vs non-typical, and discusses how the results feed into selecting the appropriate baseline model for assessing upper-body performance.
- **Chapter ??** validates our claim that individuals can adapt their kinematic behavior by interacting with our *Super Pop VRTM* system.
- **Chapter ??** concludes the dissertation by summarizing the contributions and suggestions for future work.

CHAPTER II

PREVIOUS WORK

2.1 Physical Therapy

Parkinson’s disease, stroke, and cerebral palsy are common disorders that can affect an individual’s motor skills. Parkinson’s disease is a neurodegenerative disorder that affects movement, muscle control, and balance [?, ?]. Among other side effects, people with stroke can suffer from hemiplegia (paralysis or severe weakness on one side of the body), and spasticity (stiff muscles that make limb movements difficult or uncontrollable) [?]. The term cerebral palsy (CP) describes a group of disorders of the development of movement and posture, causing activity limitations that are attributed to non-progressive disturbances in the developing fetal or infant brain [?]. Some patients may find that the difficulties of these disorders increase as they grow older and the symptoms become more severe over time. The development of these difficulties is even more common in individuals who don’t receive appropriate care. As such, physical therapy interventions are recommended as part of the rehabilitation protocol to reduce further development of the mentioned symptoms.

According to the World Confederation for Physical Therapy, “physical therapy provides services to individuals to develop, maintain, and restore maximum movement and functional ability throughout the lifespan” [?]. In general, **physical therapy** helps individuals manage their symptoms and promotes mobility and function through physical intervention. Previous studies have shown the benefits of physical therapy for individuals who have some form of motor skills disorder. For example, P. Langhorne et al. [?] reviewed randomized trials where physiotherapy was provided to individuals after stroke. The authors concluded that more intense physical therapy

following stroke (i.e. a greater amount of therapy time per day), may enhance the rate of recovery and reduce the rate of deterioration. In a study with 114 stroke patients, J. Lehmann et al. [?] showed that there were significant functional gains after their rehabilitation, which could not be attributed to spontaneous recovery. M. Morris et al. [?] argues that it's important to provide physiotherapy in conjunction with medication to individuals who have Parkinson's disease because, given that the effects of the currently available medicine can be unpredictable, physiotherapy attempts to teach these individuals how to move more easily when motor fluctuations occur. H. Anttila et al. [?] assessed the effectiveness of physical therapy interventions on motor function in children who have CP. They reviewed 22 relevant randomized clinical trials published between 1990 and 2007. Although there were some studies with low methodological quality and/or with statistically insignificant results, the team concludes that there is moderate evidence to support the effectiveness of some upper extremity intervention categories. Namely, evidence was established for the effectiveness of: upper-extremity treatments on attained goals and active supination, and constrained-induced therapy on amount and quality of hand use.

This research focuses on rehabilitation activities for children who have cerebral palsy (CP). The Center for Disease Control and Prevention reports that CP is the most common motor disability in childhood. They report that an average of 1 in 323 children in the U.S. have CP [?]. Often, individuals who have CP are unable to control some of their muscles resulting in poor movement coordination. Approximately half of the children who have CP may sustain dysfunctions in upper extremity activities such as reaching, grasping, and/or manipulating objects [?], which are the basis for performing most activities of daily living (ADLs). There are six basic ADLs that people tend to do everyday without needing assistance: eating, bathing, dressing, toileting, transferring (walking), and grooming (personal hygiene). The quality of life of children who have CP may be reduced if they need assistance to perform one or

more of these activities. As such, for individuals who have CP, especially if detected at an early age, therapists recommend participation in physical therapy interventions to reduce further development of the effects of the disorder and potentially increase their quality of life by avoiding the need of assistance to perform the ADLs.

2.2 Virtual Reality Gaming Systems

Although the importance and benefits of repeated practice of functional activities in various contexts have been emphasized [?, ?], a non-trivial amount of individuals undergoing physiotherapy interventions do not comply with the recommended physical exercises. M. Shaughnessy et al. [?] analyzed the responses of a survey of exercise beliefs and patterns. Responses from 312 older adult stroke survivors indicated that only 31% exercised four times a week. Similarly, R. Forkan et al. [?] developed a survey to determine the participation of 556 older adults in a home exercise program. Ninety percent of the participants reported receiving a home exercise program, while 37% no longer performed it. In a study to determine which factors are related to compliance with performing the recommended in-home exercises of physiotherapy protocols, E. Sluijs et al. [?] found that forgetting to exercise and lack of motivation were most often mentioned by patients who were noncompliant. The role of motivation in motor rehabilitation has been deemed to be extremely important and sometimes the most critical factor in motor rehabilitation of certain populations [?]. As such, there has been a need to develop solutions that can increase the motivation and engagement of individuals to perform the recommended in-home therapy exercises.

A recent example includes the use of virtual reality gaming systems, which have been developed to aid rehabilitation specialists in physical therapy treatments for individuals with some form of motor skill disorder. **Virtual reality** (VR) refers to a computer technology that creates a three-dimensional (3D) virtual context and virtual objects that allow for interactions by the user [?]. Therapists and researchers

have studied and shown the benefits of using VR environments as part of corresponding physical therapy interventions for various patient demographics. For example, A. Merians et al. [?] showed that feedback received during standard physical rehabilitation in combination with a VR system may lead to improvements of daily motor functions. C. Bryanton et al. [?] suggests that using VR to guide exercise may improve exercise compliance and enhance exercise effectiveness as compared to rehabilitation conventional exercises. D. Reid et al. [?] conducted a pilot study to show the benefits of a VR system for children who have CP. The study suggests that a virtual environment allows for increased play engagement and the opportunity for children to practice control over their movements. In general, VR systems that provide challenging/competitive environments may increase the user's motivation to perform their therapy exercises [?].

J. Deutsch et al. [?] used a Wii console to augment the rehabilitation of an adolescent with CP. The participant used a Wii controller to manipulate objects in the virtual environment. Although there were positive outcomes at the impairment and functional levels, the research showed that the system is limited only to individuals who are able to grasp the controller, which is not necessarily the case for all the population who has CP. M. Jannink et al. [?] used a commercially available motion capturing product for the PlayStation 2 platform called the EyeToy: Play. The team concluded that the system is not only an effective motivational training tool for the training of children with CP, but it also has the potential to improve upper extremity function - based on the two children that improved considerably relative to their Melbourne Assessment scores. Y. Chen et al. [?] also evaluated the feasibility of using the games in the EyeToy: Play for rehabilitation purposes, and concluded that they can be used to train the arm functions of children who have CP. However, given that this gaming system was not specifically developed for rehabilitation purposes, it does not allow for the individualization of the game settings to match the user's

needs.

To address this issue, multiple researchers have designed and developed their own VR games specifically for rehabilitation, many of which use the KinectTM camera from Microsoft as its motion capture system [?, ?, ?]. This is because the Kinect offers a low-cost portable solution that does not require the use of any additional equipment. However, although these systems allow for the individualization of the game settings and can be tailored to the user’s needs, they do not employ an objective and quantifiable assessment methodology to keep track of the user’s kinematic performance, which is the first step towards providing the necessary feedback users need to efficiently correct their kinematic behavior.

2.3 Upper-Body Assessment Methodologies

2.3.1 Current Clinical Methodologies

There are several assessment methodologies that physiotherapists use to keep track of their patients’ performance and improvement in their functionality as they engage in their rehabilitation protocols. All assessments measure different aspects of activity. K. Klingels et al. [?] conducted a study to identify, analyze, and compare feasible assessment methodologies of arm activities in children who have CP. Some of the methodologies identified by the authors were the Bruininks-Oseretsky Test of Motor Proficiency Second Edition (BOT-2) [?], the Peabody Developmental Motor Scale Second Edition (PDMS) [?], and the Pediatric Motor Activity Log (PMAL) [?]. The former two are used to test what children *can do* when asked to, while the latter is used to measure what children *actually do* in daily life. The BOT-2 and PDMS assess proficiency in fine and gross motor control in individuals between 4 and 21 years of age, and in children from birth through 5 years of age respectively. The PMAL provides a rating by parents about their child’s affected arm on 22 arm-hand real-world function activities (e.g. holding a cup, taking off their shoes, and turning

a knob). Parents indicate “how often” their child used their more affected hand for each activity and “how well” their child performed each activity. However, the BOT-2 and PDMS give little to no information concerning the quality of movement of the affected arm and hand, and the PMAL relies on the parent’s observations and subjective scoring, which can lead to biased results.

A meta-analysis on the use of virtual reality to treat arm function in children with CP made by Y. Chen et al. [?], revealed that studies that measured reaching kinematics with motion capture systems showed larger effect sizes than studies which used standardized clinical assessment tools. This is because measuring reaching kinematics is a more reliable approach that captures the children’s best capacity with better accuracy. As such, we adhere to this study’s findings and developed our own assessment methodology based on an individual’s kinematic performance such that the system can accurately keep track of the user’s kinematic performance.

2.3.2 Kinematic Assessment using Robot-aided Rehabilitation Systems

Previous research has shown the positive use of robot-aided rehabilitation systems in objectively assessing human movement in therapy scenarios ranging from stroke rehabilitation [?, ?, ?], to motor development in children [?]. Krebs et al. [?] presented an approach to analyze kinematic data collected using a prototype robot-aided rehabilitation facility. Not only did they show that robot-aided therapy presents no adverse effects on patients, but also that the combination of robotics and automation with concepts of neuroscience has the potential to allow for the development of better kinematic assessment tools. Volpe et al. [?] showed that individuals who acquired additional sensorimotor training delivered by a robotic device demonstrate improved functional outcome and enhanced motor performance. Clinical trials developed by Burgar et al. [?] showed that the motor recovery improvements achieved by individuals who participate in robot-assisted therapy sessions exceed the improvements

achieved by those who participate in traditional therapy sessions. In [?], Galloway et al. studied the possibility of using mobile robots as part of the rehabilitation of young infants with special needs. Finally, Colombo et al. [?] developed a robotic device to fulfill the need to assess the performance of an individual through repeatable and quantifiable metrics as an effective means for rehabilitation. Although robot-aided rehabilitation systems have shown viable use in various rehabilitation scenarios, contact-based methodologies such as these are not practical options for in-home assessment due to hardware limitations and similar restrictions. Most of these systems have also yet to be widely adopted in the clinical setting.

2.3.3 Kinematic Assessment using Non-contact Based Methods

Researchers have looked at non-contact based methods for kinematic assessment. Howard et al. [?] compared the benefits of non-contact versus contact sensing methodologies for in-home rehabilitation. They concluded that, primarily due to cost and the complexity of the system, it would be challenging to integrate the required hardware in most real-world human settings for the contact-based approach.

Butler et al. [?] developed a quantitative method to assess upper-body kinematics in children with cerebral palsy using 3D motion analysis. The method computed the participants' joint kinematics for eight primary motions of the trunk and dominant arm. However, this research is limited to assessing the participants' kinematics during a Reach and Grasp Cycle. Moreover, similar to several other studies [?, ?], this one uses a marker-based motion-capture system like the OptiTrack and Vicon systems. Due to their cost and lack of portability, such methodologies are limited to the clinical setting. Grimshaw et al. [?] used two gen-lock cameras to examine the kinematics of typical walking gait in young children. Variables analyzed included knee angles, hip vertical displacement, hip angles, among others. Finally, Brooks et al. [?] developed an objective and quantifiable methodology for assessing upper-body movements using

a portable sensing method and computer vision techniques such as Motion History Imaging, edge detection, and Random Sample Consensus. However, the method requires the user’s sagittal plane to be perpendicular to the camera’s focal vector in order for the algorithm to capture the needed images. The method also requires for the user’s elbow joint to be locked throughout any movement made. Unfortunately, these assumptions are usually not met in scenarios where the user is required to interact with a given system, for example a virtual gaming platform.

2.4 Targeted Corrective Feedback

In general, **corrective feedback** is defined as information provided to an individual during or after performing a task, relative to the quality of the task performed. Previous studies have shown that, as part of physiotherapy protocols, external feedback of performance is essential to motor learning [?]. During training, patients use feedback to detect errors in their performance by comparing their performance to the expected goal, in order to improve in the next attempt [?]. This is especially true for children who have cerebral palsy since they have limited reference to judge “good” movements from “poor” movements, and thus they must rely on external feedback to improve their performance. Moreover, it has also been shown that feedback can facilitate sustained or complex play [?], and that automatic concurrent feedback is desirable over feedback from a therapist because it’s objective and is provided faster than a human [?, ?].

Various systems for rehabilitation purposes have been developed taking these findings into consideration. K. Wood et al. [?] developed a robotic therapy system to provide feedback for specific therapeutically desirable upper-body movements. Results suggest that use of the robotic feedback system result in slightly increased movement in the targeted gesture. Authors conclude that of the reasons for these gains in motor tasks could be the immediacy of the feedback from the robotic playmate. Z. Zheng

et al. [?] developed a closed-loop autonomous robotic system that offers dynamic, adaptive, and autonomous interaction for learning of imitation skills with real-time performance evaluation and feedback. Results show that the produced performances were relatively better than that of a human therapist. R. Tang et al. [?] designed a prototype that guides users through pre-recorded physiotherapy exercises using real-time visual guides. The study concludes that participants were most accurate in performing the required tasks when using the visual guides. However, the proposed system lacks an objective assessment of the user’s kinematic performance as it focuses on evaluating the how closely the user is able to follow a given exercise. Moreover, to the best of our knowledge, none of the existing systems provide targeted corrective feedback relative to the user’s kinematic performance.

2.5 Summary of Findings

Even though it has been shown that physical therapy helps individuals with some form of motor skills disorder (e.g. Parkinson’s disease, stroke, and cerebral palsy) manage their symptoms and improve functional mobility (Section ??), there is a non-trivial amount of patients that do not comply with their recommended in-home exercises. Given that non-compliance limits the benefits of physiotherapy protocols, researchers have developed virtual reality (VR) serious games in an attempt to increase users’ engagement and intrinsic motivation to perform their in-home exercises (Section ??). The benefits of using VR gaming systems for rehabilitation purposes have been well documented. Some include:

- Feedback received during standard physical rehabilitation in combination with a VR system may lead to improvements of daily motor functions [?].
- Using VR to guide exercise may improve exercise compliance and enhance exercise effectiveness as compared to rehabilitation conventional exercises [?].

- A virtual environment allows for increased play engagement and the opportunity for children to practice control over their movements [?].
- VR systems that provide challenging/competitive environments may increase the user’s motivation to perform their therapy exercises [?].

This literature found multiple research work done in developing such types of VR systems (Section ??). However, none were found that employ an objective and quantifiable assessment methodology to keep track of the user’s kinematic performance, which is the first step towards providing the necessary feedback users need to efficiently correct their kinematic behavior. As described in Section ??, there are some limitations to current clinical upper-body assessment methodologies. The BOT-2 and PDMS methodologies give little to no information concerning the quality of movement of the affected arm and hand, and the PMAL relies on the parent’s observations and subjective scoring, which can lead to biased results. Moreover, the contact-based kinematic assessment methodologies described in Section ?? are not practical options for the in-home setting due to hardware limitations and similar restrictions, and the non-contact based methods described in Section ?? require certain assumptions to be met in scenarios where the user is required to interact with a given system, for example a virtual gaming platform. These findings motivated the need to develop our own upper-body kinematic assessment methodology. Namely, an objective method with no hardware limitations or user requirements.

Finally, Section ?? describes the importance of receiving targeted corrective feedback during physiotherapy intervention protocols. Previous studies have shown that: external feedback of performance is essential to motor learning, feedback can facilitate sustained or complex play, and that automatic concurrent feedback is desirable over feedback from a therapist because of its objectivity. The section concludes with a summary of the benefits obtained from developed systems that provide targeted

corrective feedback. As such, this dissertation takes into consideration all findings from this chapter towards the design, development, and validation of an in-home VR serious gaming system coupled with an objective and quantifiable methodology that is feasible and effective at inducing changes in users' kinematic behavior in real-time.

CHAPTER III

VIRTUAL REALITY GAMING PLATFORM

3.1 Introduction

This research presents a new in-home virtual reality (VR) game designed to improve users' upper-body mobility through the repetition of movements associated with an individualized intervention protocol. The ***Super Pop VRTM*** game [?, ?], does so mainly by allowing for the individualization of its game parameters, and by integrating clinical assessment of outcome measures as an automated objective of the system (Chapter ??).

In this chapter we discuss: 1) the ability of the *Super Pop VRTM* system to individualize the game parameters to better tailor the physiotherapy protocols to the user's needs (Section ??), 2) the filtering methods used to improve the tracking and analysis of the user's movements (Section ??), and 3) the qualitative results of a study done to measure the level of interest and enjoyment users have when interacting with the system (Section ??).

3.2 Description of Overall System

A 3D depth camera is used to track and store the user's upper-body joint coordinates during game play. This research uses the Microsoft Kinect 3D camera because of its portability, cost, performance, and ability to track and capture the coordinates of the user's joints. During game play, the user is immersed in a virtual environment where yellow, green, and red virtual bubbles appear on the screen. The goal is to 'pop' as many bubbles as possible in a certain amount of time by moving a hand over the 'good' bubbles (yellow and green) while avoiding the 'bad' bubbles (red). The 3D

depth camera keeps track of the coordinates of the user’s hands, and continuously compares them against the coordinates of the virtual bubbles shown on screen (Figure ??). In the context of the Super Pop game, ‘popping’ a bubble refers to the event when one of the user’s hands is inside the area of the corresponding bubble (??).

$$\|H_C - B_C\| \leq B_r \quad (1)$$

where H_C and B_C are the centroid coordinates of the user’s hand and bubble respectively, and B_r is the radius of the bubble.

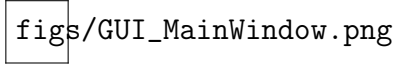


Figure 2: Main GUI of the *Super Pop VRTM* game.

The green bubbles are called **Super Bubbles** (SBs). Based on the user’s intervention protocol, there is a point in time where all yellow and red bubbles on screen are cleared and a set of two or three SBs appear one at a time. Each set of SBs highlights the trajectory that the therapist will use to evaluate the user’s rehabilitation outcome metrics. For example, if the experimental protocol is designed to improve the user’s maximum range of motion (ROM), the therapist would position three SBs such that they are spaced with a slightly greater angle than the user’s effective ROM. A 90° trajectory example is shown in Figure ?. This way, through practicing the specified repetitive motion that will appear throughout the game, the user will progressively increase his/her ROM. When ‘popping’ the SBs, the system stores the joint coordinates of the assessment arm, which is later ran through the assessment algorithm for evaluation (Chapter ?). More details on the VR system are provided in [?, ?], and the instructions on how to run the game are attached to Appendix ?.

figs/90deg_sequence.png

Figure 3: Example of a 90° trajectory created by the position of the three Super Bubbles.

3.3 *Individualization of Game Settings*

In general, physical therapists assess people who have a motor skills disorder individually because no two individuals have the same symptoms and needs. M. Morris [?] states that the goals set for a physical rehabilitation protocol are constantly modified based on the individual’s progress. Thus, we designed the *Super Pop VRTM* game such that therapists can select the most effective combination of settings that best suit their patients.

The difficulty of each game is set by selecting different combinations of the following parameters: game duration, total number of levels, game speed (rate at which the bubbles appear on screen), bad bubble ratio, bubble size, and good and bad bubble scores. These parameters serve different purposes in the rehabilitation protocols. For example, the game speed is linked to the speed of the user’s movements. Intervention protocols for users with slower movements would include games with bubbles that appear at a slower rate. The size of the bubbles and the ‘bad’ bubble ratio parameters are linked to the user’s accuracy and fine motor skills. Intervention protocols designed for users with poor accuracy and/or poor fine motor skills would include larger bubbles and/or a lower ‘bad’ bubble ratio such that the user doesn’t have to worry about avoiding bubbles. In the system’s virtual environment, a ‘bad’ bubble is a virtual object that users must avoid. This is akin to avoiding obstacles during regular physical therapy sessions. The purpose is to practice coordinated sequences towards improving the user’s fine motor control [?].

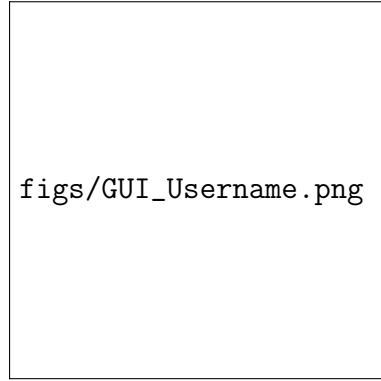
3.3.1 Game Interfaces

The game settings are selected by interacting with the system via its interfaces. The game's main graphical user interface (GUI) (Figure ??) has four buttons which allow for direct interaction with the game's interface: 'Select Username', 'Game Settings', 'Bubble Appearance Region', and 'Start/Restart Super Pop Game'. The first three buttons access secondary GUIs that provide the therapist options for customizing the intervention protocol of the game (Figure ??). The 'Select Username' GUI lets the therapist assign individual usernames or IDs to the users allowing for the system to keep track of multiple players at different times (Figure ??). The 'Game Settings' GUI offers the option to choose from three different game difficulties with pre-selected parameters (i.e. Easy, Normal, and Hard), as well as a Custom option (Figure ??). The Custom option enables the therapist to provide their own combination of game settings relative to the user's needs. Moreover, therapists can also select the shape of the 'good' and 'bad' bubbles (circles, squares, or triangles), and they can select the notes of the song that will play whenever a bubble is 'popped'. This interface also provides several secondary options that enhance game performance. For example, if a user plays the game while seated, the therapist can improve the camera's tracking accuracy by selecting the appropriate tracking mode. The 'Bubble Appearance Region' GUI shows a snapshot of the user in the virtual environment (Figure ??). In this interface, the therapist can select the Super Bubble (SB) appearance interval duration, the number of SBs used for the protocol, and select the arm to be assessed. Moreover, the interface also allows for personalized sessions accommodating the different body structures of the users. It enables the therapist to select the workable region in which regular bubbles will appear and the position of the SBs based on the user's placement relative to the camera.

3.3.2 Evaluation Trajectory

In general, therapists may prompt their patients to follow a path during a physiotherapy session to analyze how well they can control their movements [?]. One way the *Super Pop VRTM* game allows for personalized sessions accommodating different body structures and intervention protocols is by defining an evaluation trajectory that corresponds to the user’s needs. In the Super Pop environment, an evaluation trajectory is defined by selecting the positions of the SBs, which can be selected by one of three available methods: 1) double-clicking the positions of each SB in the ‘Bubble Appearance Region’ interface, 2) specifying the exact (x, y) pixel coordinates, or 3) specifying a trajectory by its angle of separation. The latter two options are available to the user by accessing the ‘Set Coordinates’ GUI (Figure ??).

Specifying a trajectory by its angle of separation refers to positioning the SBs in a path that overlays the outline of a circle centered at the shoulder of the user’s assessment arm. The bubbles are equally separated from each other based on the amount of degrees selected. An example of a 90° trajectory is shown in Figure ?. Before the game starts the therapist can ask the user to raise his/her arms as high as possible such that the SBs are positioned relative to the user’s maximum reach. When the therapist clicks on the ‘Bubble Appearance Region’ button, the system defines a circle centered on the shoulder of the user’s assessment arm, in which the radius is the Euclidean distance between the shoulder and hand of the user’s assessment arm. This radius is not necessarily equal to the length of the user’s arm. For instance, a child who has some form of motor skills disorder may not be able to extend his/her arm to its full length. The purpose is to encourage users to practice movements which require them to extend their arms as much as they can such that they can regain that ability. The first SB is positioned at the same coordinates as those of the hand of the user’s assessment arm. The other two SBs are positioned according to the description in Algorithm ?, which follows the labels shown in Figure ?.



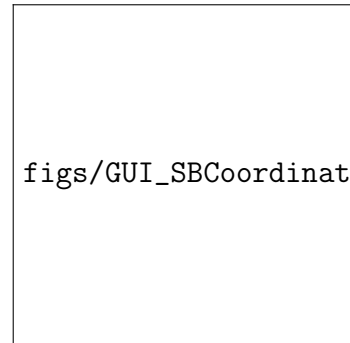
(a) Select Username



(b) Game Settings



(c) Bubble Appearance Region



(d) Set SB Coordinates

Figure 4: *Super Pop VRTM* game's secondary GUIs.

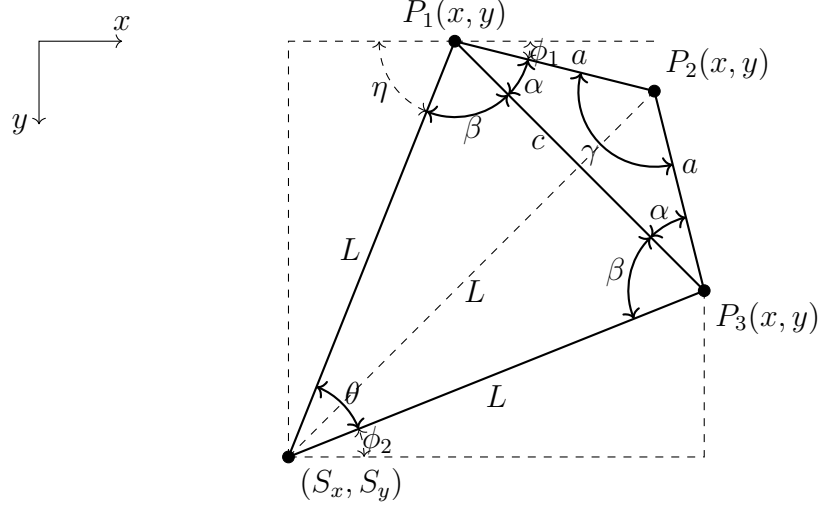


Figure 5: Diagram with link and angle labels for computing the SBs' coordinates, where variables are described in Algorithm ??.

3.4 Filtering Methods

To reduce noise in the data collection process of joint coordinates, we implemented two different filtering methods. The first filter is readily available as part of the **Microsoft Kinect's SDK** (software development kit). It is applied during game play to improve the accuracy of the camera's tracking capabilities. The second filter is the **Moving Average Filter**. It is applied to the collected data in the post-processing stage to reduce the effect of potential outliers. It is important that the trajectories that are generated with the captured upper-body joint coordinates are relatively smooth such that they are accurate enough when using them to analyze the user's kinematic performance. In general, a moving average filter is a one-dimensional filter that analyzes one element at a time from a given data set. For each element in the data set, the filter computes an average value based on the neighboring elements and replaces the analyzed element with the computed average. The method through which the filter smooths the i^{th} element of the original data is described in (??). This process is repeated for all the elements in the given data set, except for the first and last points which remain the same.

Algorithm 1 Compute the (x, y) pixel coordinates of the three SBs

Require: L =distance between the shoulder and hand of the user's assessment arm, (H_x, H_y) =coordinates of the user's hand, (S_x, S_y) =coordinates of the user's shoulder, θ =angle of separation between the three SBs, and **arm**=assessment arm

```

1:  $\begin{bmatrix} P_{1x} \\ P_{1y} \end{bmatrix} \leftarrow \begin{bmatrix} H_x \\ H_y \end{bmatrix}$   $\triangleright \begin{bmatrix} P_{1x} \\ P_{1y} \end{bmatrix}$ :  $(x, y)$  coordinates of SB1
2:  $a \leftarrow L\sqrt{2(1 - \cos(\theta/2))}$ 
3:  $c \leftarrow L\sqrt{2(1 - \cos(\theta))}$ 
4:  $\gamma \leftarrow \cos^{-1}[(2a^2 - c^2)/(2a^2)]$ 
5:  $\alpha \leftarrow (1/2)(\pi - \gamma)$ 
6:  $\beta \leftarrow (1/2)(\pi - \theta)$ 
7:  $\eta \leftarrow \cos^{-1}[(1/L)(P_{1x} - S_x)]$ 
8:  $\phi_1 \leftarrow \pi - (\alpha + \beta + \eta)$ 
9:  $\phi_2 \leftarrow \theta - \eta$ 
10:  $\begin{bmatrix} P_{2x} \\ P_{2y} \end{bmatrix} \leftarrow \begin{bmatrix} P_{1x} + a \cos(\phi_1) \\ P_{1y} + a \sin(\phi_1) \end{bmatrix}$   $\triangleright \begin{bmatrix} P_{2x} \\ P_{2y} \end{bmatrix}$ :  $(x, y)$  coordinates of SB2
11:  $\begin{bmatrix} P_{3x} \\ P_{3y} \end{bmatrix} \leftarrow \begin{bmatrix} S_x + L \cos(\phi_2) \\ S_y + L \sin(\phi_2) \end{bmatrix}$   $\triangleright \begin{bmatrix} P_{3x} \\ P_{3y} \end{bmatrix}$ :  $(x, y)$  coordinates of SB3
12: if  $arm = Left$  then
13:    $\begin{bmatrix} P_{2x} \\ P_{2y} \end{bmatrix} \leftarrow R(-\theta) \begin{bmatrix} P_{2x} - S_x \\ P_{2y} - S_y \end{bmatrix} + \begin{bmatrix} S_x \\ S_y \end{bmatrix}$   $\triangleright R(-\theta)$ : rotation matrix around  $-\theta$ 
14:    $\begin{bmatrix} P_{3x} \\ P_{3y} \end{bmatrix} \leftarrow R(-2\theta) \begin{bmatrix} P_{3x} - S_x \\ P_{3y} - S_y \end{bmatrix} + \begin{bmatrix} S_x \\ S_y \end{bmatrix}$   $\triangleright R(-2\theta)$ : rotation matrix around  $-2\theta$ 
15: end if
16: return  $(P_{1x}, P_{1y}), (P_{2x}, P_{2y}), (P_{3x}, P_{3y})$ 

```

$$Y_S(i) = \begin{cases} \frac{1}{2N+1} \sum_{j=-N}^N Y(i+j), & \text{if } i = [2, m-1] \\ Y(i), & \text{otherwise} \end{cases} \quad (2)$$

where $Y_S(i)$ is the i^{th} point in the new smoothed data set, $Y(i)$ is the i^{th} point in the original data set, m is the total number of points in the data set, and N is the window size of the filter. The window size N is empirically selected based on how well the span $2N+1$ works on the collected data points. We compared three different window sizes on different joints and axes (Figure ??). We concluded that a window size of $N = 2$ yields the best results when compared to sizes of $N = 3$ and $N = 4$ because it smooths the original data without losing potential important information

as the other two window sizes seem to do.

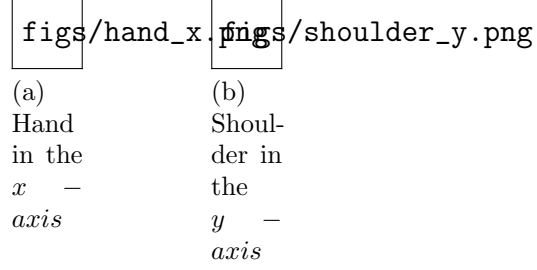


Figure 6: Comparison of three window sizes for the Moving Average Filter.

3.5 Engagement Study

3.5.1 Hypotheses

One of the *Super Pop VRTM* game's purposes is to increase the user's intrinsic motivation to perform the recommended in-home physical therapy exercises. We conducted a study to determine the level of users' interest/enjoyment when interacting with the system to show that the game can indeed increase its users' intrinsic motivation [?]. The following were the hypotheses for this study:

H_1 : The *Super Pop VRTM* game maintains its users' interest and enjoyment.

H_2 : Users experience minimal pressure and tension when interacting with the *Super Pop VRTM* game.

3.5.2 Experimental Design

Fourteen able-bodied adults were recruited to interact with the *Super Pop VRTM* system. Nine females and five males ranging in age between 18 and 31 years played the game. All participants signed the IRB (Institutional Review Board) approved consent form allowing them to participate in the testing sessions. We recruited volunteers who had never had any previous interactions with the system to eliminate the bias that their relative experience might provide. Participants were asked to play two rounds

of the game such that they had enough time to get familiarized with the system. After play, they were asked to choose the most appropriate responses to a subset of selected statements randomly ordered from the 7-point Likert scale that is the **Intrinsic Motivation Inventory (IMI)**; where 1 means ‘not at all true’, 4 means ‘somewhat true’, and 7 means ‘very true’. The IMI is a device intended to assess participants’ subjective experience related to a target activity in laboratory experiments [?]. The instrument measures the participants’ self-reported motivation by assessing several sub-scales: interest/enjoyment, perceived competence, effort/importance, pressure/tension, and value/usefulness. The interest/enjoyment scale is considered the self-report measure of intrinsic motivation, thus we focus the better part of the analysis on this sub-scale. The definition for the other sub-scales can be found in [?], and the survey used in this study is attached to Appendix ??.

To validate the participants’ responses, we performed independent *t*-tests on each sub-scale. The purpose was to determine if the participants’ responses are statistically significant relative to the purpose of each sub-scale. This is to say, we hope to achieve responses close to the ‘very true’ statement (i.e. score of 7) for the sub-scales with positive predictors of self-report motivation, and responses close to the ‘not at all true’ statement (i.e. score of 1) for the sub-scales with negative predictors of self-report motivation. As such, for each sub-scale, we performed a *t*-test with respect to each of the seven score values of the Likert scale. All tests were performed at a 99% confidence level and are of the null-hypothesis that, for each of the seven score values, the responses from the given sub-scale come from a distribution with mean equal to the corresponding score value. The value at which the null-hypothesis can’t be rejected reflects the best approximation of the responses’ real mean.

3.5.3 Results

We compiled the answers of all participants and organized the scores relative to the selected IMI sub-scales. The average IMI scores ± 1 std for each sub-scale are shown in Figure ???. For each sub-scale, the p -values for each score in the 7-point Likert scale are shown in Table ??.

figs/all_scores_nocomparison_updatedSTD2.png

Figure 7: Average IMI scores from all participants organized by sub-scale.

Table 1: p -values for each sub-scale relative to each of the 7-point Likert scale scores at a 99% confidence level.

Sub-scales	Scores in the 7-point Likert scale						
	1	2	3	4	5	6	7
Effort/Importance	$\ll 0.01$	$\ll 0.01$	$\ll 0.01$	< 0.01	0.359	$\ll 0.01$	$\ll 0.01$
Interest/Enjoyment	$\ll 0.01$	$\ll 0.01$	$\ll 0.01$	$\ll 0.01$	$\ll 0.01$	0.047	$\ll 0.01$
Perceived Competence	$\ll 0.01$	$\ll 0.01$	$\ll 0.01$	$\ll 0.01$	$\ll 0.01$	0.898	$\ll 0.01$
Pressure/Tension	$\ll 0.01$	0.541	$\ll 0.01$	$\ll 0.01$	$\ll 0.01$	$\ll 0.01$	$\ll 0.01$
Value/Usefulness	$\ll 0.01$	$\ll 0.01$	$\ll 0.01$	$\ll 0.01$	< 0.01	< 0.01	$\ll 0.01$

*Bold p -values are the ones for which the null-hypothesis can't be rejected at a 99% confidence level.

3.5.4 Discussion and Conclusions

The interest/enjoyment sub-scale is the one that mainly assesses the participants' intrinsic motivation. Out of a maximum score of 7.0, the average score from all participants was **5.7 \pm 1.30** (Figure ??). From Table ??, the null-hypothesis is rejected for all the scores below the neutral score of 4, but not for the score of 6 ($p = 0.047$). This suggests that it is highly likely the average response from all participants is on the 'true' side of the scale. The validation of the relatively high score for the interest/enjoyment sub-scale supports our claim that the *Super Pop VRTM* game is perceived as engaging and interesting thus validating our first hypothesis (H_1). Moreover, the score for the pressure/tension sub-scale was relatively low (**2.1 \pm 1.56**). The null-hypothesis is rejected for all the scores above the neutral score

of 4, but not for the score of 2 ($p = 0.541$). This suggests that it is highly likely the average response from all participants in on the ‘not true’ side of the scale. The validation of the relatively low score for the pressure/tension sub-scale supports our claim that, in general, the *Super Pop VRTM* game promotes a relatively relaxing environment where users can perform their therapy exercises with minimal pressure or tension thus validating our second hypothesis (H_2).

These results correlate with the remaining sub-scales as well. The average responses for the perceived competence and value/usefulness sub-scales were relatively high, 6.0 ± 0.92 and 5.5 ± 1.11 respectively. The perceived competence sub-scale falls in the ‘true’ side of the scale given that the null-hypothesis can’t be rejected for the score of 6, as shown in Table ?? ($p = 0.898$). We performed a separate t -test on the value/usefulness sub-scale since the null-hypothesis is rejected for all seven scores. The test was also at a 99% confidence level and of the null-hypothesis that the responses in the sub-scale come from a distribution with mean equal to 5.5. With a p -value of $p = 0.831$, the null-hypothesis is not rejected for a score greater than the neutral score of 4, thus the value/usefulness sub-scale also falls in the ‘true’ side of the scale. These results suggest that, up to some extent, participants identified having a sense of competence in their actions while interacting with the *Super Pop VRTM* game, and that the Super Pop system is useful in the context of rehabilitation.

This study’s results are limited by the targeted population. The volunteers that were recruited for this study were young adults versus children with a motor skill disorder. Another item to keep in mind is that there is a possibility that the users’ self-reported motivation may vary as a function of their experience with the system. In the meantime, we conclude that the developed system has the potential to be useful in the rehabilitation setting. By maintaining the user’s interest and enjoyment while reducing the pressure and tension of complying with the recommended intervention protocol, the *Super Pop VRTM* game can increase the efficacy of said protocol.

3.6 Summary

This chapter described the virtual reality (VR) serious game we developed and its potential as an in-home rehabilitation tool for individuals who have a motor skill disorder. Namely, the chapter describes the game’s ability to individualize its game settings such that it can tailor the gaming experience to the user’s needs. One example as to how the game allows for individualized sessions is by employing an algorithm that selects an appropriate evaluation trajectory as a function of the user’s upper-body dimensions and functional capabilities (Algorithm ??). This allows for a system that prompts users to complete specific reaching tasks in the virtual environment, leading them towards making progress in their intervention protocols.

To address the problem of individuals undergoing physiotherapy interventions not complying with the recommended exercises [?], the second part of the chapter described the engagement study we conducted. The objective was to determine how well the *Super Pop VRTM* game maintains the user’s engagement and interest with the ultimate goal of increasing the patients’ motivation to comply with their intervention protocols [?]. We employed the Intrinsic Motivation Inventory (IMI) to measure the participants’ self-reported motivation by assessing several sub-scales like interest/enjoyment and pressure/tension. The study concludes that, in general, the *Super Pop VRTM* game promotes a relatively relaxing environment where users can perform their therapy exercises with minimal pressure or tension while having a sense of competence in their actions.

CHAPTER IV

OBJECTIVE AND QUANTIFIABLE ASSESSMENT METHODOLOGY

4.1 Introduction

In the rehabilitation field, determining the effectiveness of an intervention protocol begins by comparing the individual’s movement characteristics against a baseline. In most settings, this baseline is determined through clinical studies involving a range of patients belonging to the same demographic group [?, ?, ?, ?]. Unfortunately, this leads to a process that is difficult to repeat for all patient demographics or all movement characteristics given the demands on clinicians’ and patients’ time for performing such clinical baseline measurement studies. For example, Butler et al. [?] compared the data from two children with moderate spastic hemiplegic cerebral palsy to the data obtained from 25 typically developing children and adolescents. Although useful results, it’s not a trivial task to recruit 25 typically-developing children in a timely fashion to conduct the experimental sessions.

Instead of physically collecting human data, we propose a kinematic model that generates a baseline in real-time for different kinematic parameters relative to the state of a given task. This allows for an adaptable baseline that is individualized to each user’s upper-body dimensions and movement characteristics. Such a system would also enable evaluation for in-home rehabilitation, which would assist in evaluating intervention protocols outside of the clinical setting. We make this proposition under the assumption that a mathematical model of human movement will yield the most efficient results in terms of the kinematic parameters selected to describe the movements of interest. The kinematic model described in this section is validated

in Section ?? by comparing the outcome measures obtained from our model to the outcome measures obtained from a user-based baseline.

This chapter focuses on upper-body tasks, of which the most dominant form is reaching movements. The ability to reach - which has been defined as the voluntary positioning of the hand at or near a desired location [?] - is critical for most, if not all, activities of daily living such as feeding, grooming, and dressing. Moreover, failure to substantially recover upper-extremity function can lead to depression [?]. As such, reaching movements, correlated to reaching exercises, are of interest in various rehabilitation scenarios. In general, reaching movements require an individual to move from a defined initial position to a selected target position (Figure ??). In the Super Pop environment, reaching movements refer to moving from a ‘starting bubble’ to a ‘target bubble’ (Figure ??). We define a model by constructing a kinematic chain of links that correlates to the dynamics of the human arm (Section ??). A detailed description of the complete kinematic model can be found in [?].

□figs/reaching_sequence_blurred.jpg

Figure 8: Sequence of snapshots showing an example of a typical ‘reaching movement’.

4.2 *Kinematic Model of the Human Arm: Penalized Manipulator Jacobian*

A representation of this kinematic chain is shown in Figure ??, where θ_i is the angle of the i^{th} joint and L_j is the length of the j^{th} link. The model includes one spherical joint (shoulder) that provides three degrees of freedom (DOF) and one revolute joint (elbow) that provides one. For our application, the wrist joint is modeled as part of the forearm/hand link, thus resulting in a 4 DOF model.

Each transformation between joints is modeled using lie groups. Exploiting the lie group’s structure, which accounts for both the link lengths and DOF, the user’s hand position, g_w , can be estimated using forward kinematics as shown in (??).

figs/arm_kinematics.png

Figure 9: Mapping from (a) the human arm kinematics, to (b) the kinematics of a common 4 DOF robotic manipulator. (Image adapted from [?].)

$$g_w = \begin{bmatrix} R_z(\theta_1) & \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \\ 0 & 1 \end{bmatrix} * \begin{bmatrix} R_y(\theta_2) & \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \\ 0 & 1 \end{bmatrix} * \begin{bmatrix} R_x(\theta_3) & \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \\ 0 & 1 \end{bmatrix} * \begin{bmatrix} R_x(\theta_4) & \begin{bmatrix} 0 \\ 0 \\ -L_1 \end{bmatrix} \\ 0 & 1 \end{bmatrix} * \begin{bmatrix} I_{3 \times 3} & \begin{bmatrix} 0 \\ 0 \\ -L_2 \end{bmatrix} \\ 0 & 1 \end{bmatrix} \quad (3)$$

where $R_x, R_y, R_z \in SO(3)$ denote the rotation matrices along the x, y, z axis respectively, θ_i is the angle of the i^{th} joint, L_j is the length of the j^{th} link, and $I_{3 \times 3}$ is a 3x3 identity matrix. Using this kinematic model of the arm, we can solve for a trajectory from a start to a target point using the **Penalized Manipulator Jacobian** [?] as it minimizes the energy in (??) for a given reaching task.

$$E = \frac{1}{2} \|\varepsilon - J(d\theta)\|^2 + \frac{1}{2} \|d\theta^T W d\theta\|^2 \quad (4)$$

where W represents the penalization of the respective joint angles (θ), $d\theta$ is the angle update, and ε is the desired twist. This is a least squares problem and is thus a linear approximation of the complex problem. However, given a relatively small update it works very well for our problem.

The user's initial pose is estimated using the model and is set as the initial condition for the solver. For our system, we assume that only the velocity components of the twist are required. We found this to be a fair assumption as the general motions typically observed during reaching movements involve direct translations of the user's hand, with little to no wrist rotation observed. The velocity components are represented by the first two components of the ε vector, and are equated as the normalized

difference of the initial and final position of the end-effector. Solving for (??) at each iteration, $d\theta$ is computed and the joint angles are updated according to (??).

$$\theta(k+1) = \theta(k) + \delta d\theta(k) \quad (5)$$

where $\theta(k)$ is the joint angle at the k^{th} iteration, and δ is a constant from 0 to 1. This process is continued until the hand position of the model has converged to the position of the target.

The model was designed such that the movements are not only feasible, but also efficient. The penalization values in W are fixed and empirically selected depending on whether the motion would prioritize a heavily elbow or shoulder dependent motion. This prevents the model from creating trajectories that would be akin to an individual overexerting themselves. The diagonal terms of the matrix W represent the corresponding weights for their respective joints. The magnitude of each weight is greater or equal to zero, with zero representing no penalization for that given joint motion. To prioritize the motion of a given joint, the magnitude of its weight must be an order smaller than that of the joints that are not to be prioritized. For example, if the motion desired should prioritize an elbow motion (i.e. the task requires the user to bend the elbow), then the elbow weight should be much smaller than the weight of the shoulder. Similarly, for a task that restricts the user from bending his/her elbow, the shoulder weight should be much smaller than the elbow weight. The weights are empirically selected as a function of the behavior to be modeled.

Given the task’s corresponding initial and target points and the user’s arm length, the final model generates a list of waypoints that describe an efficient trajectory between the initial and target points. The resulting trajectory is a curve in \mathbb{R}^3 that matches the structure of an individual’s curve - based on the findings of [?, ?] that showed that hand paths during reaching movements are straight or slightly curved. We use the generated waypoints to define the baseline trajectory as the optimal

solution to the current task. These waypoints are then used to compute the baseline values for the kinematic parameters of interest (Section ??). An example of the comparison between the baseline’s and a user’s trajectories is shown in Figure ??.

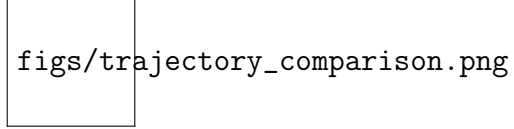


Figure 10: Comparison of the trajectory created by the user (dotted line), and the trajectory created by the kinematic model (continuous line).

4.3 *Kinematic Parameters*

It has been shown that kinematic measures during a reaching task are correlated to functional measures of upper-extremity function [?], and that kinematic measures can be used to assess performance [?]. Based on these facts, we define seven kinematic parameters that mathematically describe an individual’s upper-body movements during a reaching task. The following are general descriptions of how our algorithm computes each parameter and how it compares the user’s kinematic performance to that of the model. We also defined a percent error or error ratio equation for some of the parameters to objectively quantify the difference between the user’s and model’s performances.

4.3.1 **Deviation from Path**

In the context of a human moving their arm between two points in space, it can be shown that there exists an energy-efficient path between these two points. Using the length of the user’s arm links as a constraint, we assume that the path taken by the kinematic model is the optimal path between the two given points. As such, the algorithm compares the path taken by the user to the corresponding optimal path for a given task and computes a cost value greater or equal to zero that determines how far off the user’s movements are, spatially, from the optimal trajectory. The results

from this parameter gives us insight as to how well an individual can control their movements [?]. A large cost value suggests that the user may have troubles with motor control and, as such, is not able to follow a path correctly between two points.

In general, let $\gamma : [0, 1] \times W \rightarrow \mathbb{R}^n$ define a curve connecting N waypoints $w_i \in W \subset \prod_{i=1}^N \mathbb{R}^n$. In particular, the curve is defined as the straight line segments connecting consecutive points w_i and w_{i+1} , $\forall i = [1, \dots, N-1]$. The arc length of the curve γ , $L(\gamma)$, is then the sum of the length of the individual line segments connecting two consecutive points, given by (??),

$$L(\gamma) = \sum_{i=2}^N L_i = \sum_{i=2}^N \|w_i - w_{i-1}\| \quad (6)$$

It will be convenient to compute the arc length of the curve γ up to the individual points w_m , which we will denote using the notation (??),

$$L(w_m) = \sum_{i=2}^m L_i = \sum_{i=2}^m \|w_i - w_{i-1}\| \quad (7)$$

with $1 < m \leq N$. Notice that the total arc length of the curve can be then expressed as, $L(\gamma) = L(w_N)$. Using (??), we compute the arc length of curve γ at each of its N waypoints to create θ , a vector approximation of the arc length parameterization given by (??),

$$\theta = \begin{bmatrix} L(w_1) \\ L(w_2) \\ \vdots \\ L(w_N) \end{bmatrix} \quad (8)$$

where we use the convention $\mathbf{L}(w_1) \equiv \mathbf{0}$ and $\mathbf{max}(\theta) = \mathbf{L}(\gamma)$.

It will also be convenient to normalize the arc length parameterization approximation with respect to their corresponding maximum arc lengths to obtain $\hat{\theta}$ (??),

$$\hat{\theta} = \frac{1}{L(w_N)} \begin{bmatrix} L(w_1) \\ L(w_2) \\ \vdots \\ L(w_N) \end{bmatrix} \quad (9)$$

Notice that even when the model's and user's trajectory curves have different lengths, they can be compared after parameterizing and normalizing them since $\mathbf{max}(\hat{\theta}) = 1$.

In the case of 3D curves, we then separate each curve into their corresponding (x, y, z) axes for simplicity. The two curves from Figure ?? are shown in Figure ?? as a function of $\hat{\theta}$. These are separated into their respective axes and normalized such that their lengths are equal to 1. We then interpolate both curves in their respective axes such that both curves have the same number of waypoints, and such that the points in each curve are equally distributed with respect to their arc lengths. This allows us to compute the area between the two curves for each axis using the Right Riemann sum (an approximation of the form $\sum f(x) dx$) with $dx = 1/N$, where N is the pragmatically selected number of interpolation points.


 `trajectory_comparison_axes.png`

Figure 11: Comparison of the trajectory created by the user (dotted line), and the trajectory created by the kinematic model (continuous line) separated into their respective (x, y, z) axes as a function of their normalized arc lengths.

Finally, the Deviation from Path (DfP) parameter computes the total area between two 3D curves, whose waypoints we stack into matrices C_1 and C_2 , with (??), where A_x , A_y , and A_z are the areas in the x , y , and z axes respectively. Thus, $\text{DfP}(C_1, C_2)$ is the final cost that defines how much the user's movement deviates from the generated baseline for the corresponding task, where $\text{DfP}(C_1, C_2) = 0$ means that the two curves are spatially exactly the same.

$$\text{DfP}(C_1, C_2) = \sqrt{A_x^2 + A_y^2 + A_z^2} \quad (10)$$

4.3.1.1 DfP Parameter is a Metric

In general, a **metric** is a function that measures the distance between two vectors in space and outputs a scalar value in the set of non-negative real numbers (??).

$$D : \mathbb{R}^n \times \mathbb{R}^n \rightarrow [0, \infty) \quad (11)$$

For all \vec{x} , \vec{y} , and $\vec{z} \in \mathbb{R}^n$, $D(\vec{x}, \vec{z})$ is considered to be a metric if all the conditions in (??) are met.

$$D(\vec{x}, \vec{z}) \geq 0 \quad (12a)$$

$$D(\vec{x}, \vec{z}) = 0 \iff \vec{x} = \vec{z} \quad (12b)$$

$$D(\vec{x}, \vec{z}) = D(\vec{z}, \vec{x}) \quad (12c)$$

$$D(\vec{x}, \vec{z}) \leq D(\vec{x}, \vec{y}) + D(\vec{y}, \vec{z}) \quad (12d)$$

To show that the DfP parameter is a metric, let's first expand the equation in (??). Let $C_1, C_2 \in \mathbb{R}^{N \times 3}$ be matrices containing N interpolated waypoints that define the reparameterized and normalized 3D curves to be compared, and let $(\vec{x}_1, \vec{y}_1, \vec{z}_1)$ and $(\vec{x}_2, \vec{y}_2, \vec{z}_2) \in \mathbb{R}^N$ be the corresponding (x, y, z) components of C_1 and C_2 , respectively. Thus, we compute the areas A_x , A_y , and A_z with (??), where $dx = 1/N$.

$$\begin{aligned} A_x &= \sum_{i=2}^N \|\vec{x}_1(i) - \vec{x}_2(i)\| dx \\ A_y &= \sum_{i=2}^N \|\vec{y}_1(i) - \vec{y}_2(i)\| dx \\ A_z &= \sum_{i=2}^N \|\vec{z}_1(i) - \vec{z}_2(i)\| dx \end{aligned} \quad (13)$$

The expanded equation for the DfP parameter is then given by (??), where, for example, $\vec{x}_1(i)$ is the i^{th} element in \vec{x}_1 .

$$\begin{aligned} \text{DfP}(C_1, C_2) &= \sqrt{A_x^2 + A_y^2 + A_z^2} \\ &= \left[\left(\sum_{i=2}^N \|\vec{x}_1(i) - \vec{x}_2(i)\| \right)^2 + \left(\sum_{i=2}^N \|\vec{y}_1(i) - \vec{y}_2(i)\| \right)^2 + \left(\sum_{i=2}^N \|\vec{z}_1(i) - \vec{z}_2(i)\| \right)^2 \right]^{1/2} dx \end{aligned} \quad (14)$$

For simplicity, instead of verifying if (??) meets the conditions in (??), we can verify only one of its components. This is because (??) can be viewed as the L_2 norm of A_{total} (i.e. $\text{DfP}(C_1, C_2)$ is equal to the square root of the sum of the squares of A_{total} 's components). This, combined with the fact that the equations for the A_x , A_y , and A_z components have the same structure (??), allows us to conclude that the DfP equation meets the same conditions as one of its components. Thus, let's verify if A_x , as defined in (??), meets the necessary conditions for it to be considered a metric.

1. *Proof.* Non-negativity (??):

A_x is the sum of the distances between two points in space. By definition, distance functions map to non-negatives numbers. Squaring non-negative quantities, adding them to other non-negative quantities, and taking the square root preserves the non-negativity required. \square

2. *Proof.* Identity of Indiscernibles (??):

Let $C_1 = C_2$. This is to say that both curves have the same waypoints (i.e. $(\vec{x}_1, \vec{y}_1, \vec{z}_1) = (\vec{x}_2, \vec{y}_2, \vec{z}_2)$). Thus, A_x in $\text{DfP}(C_1, C_1)$ would yield:

$$A_x = \sum_{i=2}^N \|\vec{x}_1(i) - \vec{x}_1(i)\| = 0$$

Moreover, $A_x \neq 0$ for any other scenario where the two input curves do not have the same waypoints. As this is also the case for A_y and A_z , we obtain that $\text{DfP}(C_1, C_1) = 0$. \square

3. *Proof.* Symmetry (??):

Because the norm of a vector is a function that assigns a strictly positive value, the order of its arguments does not affect the outcome. Thus,

$$A_x = \sum_{i=2}^N \|\vec{x}_1(i) - \vec{x}_2(i)\| = \sum_{i=2}^N \|\vec{x}_2(i) - \vec{x}_1(i)\|$$

As this is also the case for A_y and A_z , we obtain that $\text{DfP}(C_1, C_2) = \text{DfP}(C_2, C_1)$. □

4. *Proof.* Triangle Inequality (??):

Let $(\vec{x}_1, \vec{y}_1, \vec{z}_1)$, $(\vec{x}_2, \vec{y}_2, \vec{z}_2)$, and $(\vec{x}_3, \vec{y}_3, \vec{z}_3) \in \mathbb{R}^n$ be the (x, y, z) components containing the waypoints of the curves C_1 , C_2 , and C_3 respectively. Analyzing the x -axis of the three curves:

$$\begin{aligned} D(\vec{x}_1, \vec{x}_3) &\leq D(\vec{x}_1, \vec{x}_2) + D(\vec{x}_2, \vec{x}_3) \\ \sum_{i=2}^N \|\vec{x}_1(i) - \vec{x}_3(i)\| &\leq \sum_{i=2}^N \|\vec{x}_1(i) - \vec{x}_2(i)\| + \sum_{i=2}^N \|\vec{x}_2(i) - \vec{x}_3(i)\| \end{aligned}$$

Taking the square yields

$$\begin{aligned} &\left(\sum_{i=2}^N \|\vec{x}_1(i) - \vec{x}_2(i)\| + \sum_{i=2}^N \|\vec{x}_2(i) - \vec{x}_3(i)\| \right)^2 \\ &= \left(\sum_{i=2}^N \|\vec{x}_1(i) - \vec{x}_2(i)\| \right)^2 + \left(\sum_{i=2}^N \|\vec{x}_2(i) - \vec{x}_3(i)\| \right)^2 \\ &\quad + 2 \left(\sum_{i=2}^N \|\vec{x}_1(i) - \vec{x}_2(i)\| \right) \left(\sum_{i=2}^N \|\vec{x}_2(i) - \vec{x}_3(i)\| \right) \end{aligned}$$

The area between any two curves will always be less or equal than the sum of the areas between each of the two curves and a third one. □

Given that A_x , as defined in (??), meets all the conditions in (??), $\text{DfP}(C1, C2)$ meets them as well. The DfP parameter is thus classified as a metric. The other parameters defined in this research do not meet all the necessary conditions.

4.3.2 Elbow and Shoulder Range of Motion

Therapists are also interested in evaluating an individual's joint flexibility (i.e. the range of motion allowed for that joint) [?]. It determines the extent of the range of movement an individual has for a specific joint. For upper-body movements, range of motion (ROM) is associated with shoulder, elbow, and wrist movements, which can then be subdivided into abduction/adduction, and flexion/extension movements. For example, for the shoulder joint, abduction/adduction is the movement in the coronal plane whereas flexion/extension movements are in the sagittal plane (Figure ??).

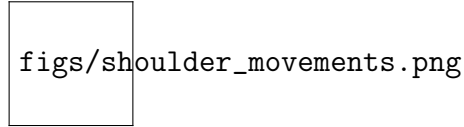


Figure 12: Examples of (a) an abduction movement in the coronal plane, and (b) a flexion movement in the sagittal plane for the shoulder joint. (Image adapted from [?]).

The literature reveals that there have been several tools designed to measure joint ROM [?]. The literature also reveals that there are yet no standard guidelines for selecting the position of the individuals to obtain the most reliable measurements [?, ?, ?]. Researchers conclude that, for the most reliable results, all clinical measures should be made in a consistent position. Thus, we define our procedures for measuring the ROM of the joints of interest relative to the nature of the *Super Pop VRTM* game. Namely, we measure the elbow's and shoulder's 3D angles given that users interact with the virtual environment by moving their arms in the 3D space with no plane restrictions.

As per the definition in [?], the effective ROM of a joint for a given movement is the absolute difference between the joint's angles at the initial and final positions of the movement. We adhere to the angle references used by therapists in different rehabilitation scenarios similar to [?] (Figure ??). The elbow ROM is a function

of the initial and final positions of the elbow-shoulder (\overrightarrow{ES}) and elbow-wrist (\overrightarrow{EW}) vectors as described by (??).

$$\begin{aligned}
ROM_E &= \left| \theta_E^f - \theta_E^i \right| \\
&= \left| \left[\pi - \cos^{-1} \left(\frac{\overrightarrow{ES}^f \cdot \overrightarrow{EW}^f}{\|\overrightarrow{ES}^f\| \|\overrightarrow{EW}^f\|} \right) \right] - \left[\pi - \cos^{-1} \left(\frac{\overrightarrow{ES}^i \cdot \overrightarrow{EW}^i}{\|\overrightarrow{ES}^i\| \|\overrightarrow{EW}^i\|} \right) \right] \right| \quad (15) \\
&= \left| \cos^{-1} \left(\frac{\overrightarrow{ES}^f \cdot \overrightarrow{EW}^f}{\|\overrightarrow{ES}^f\| \|\overrightarrow{EW}^f\|} \right) - \cos^{-1} \left(\frac{\overrightarrow{ES}^i \cdot \overrightarrow{EW}^i}{\|\overrightarrow{ES}^i\| \|\overrightarrow{EW}^i\|} \right) \right|
\end{aligned}$$

where θ_E^i and θ_E^f are the initial and final elbow angles respectively, \overrightarrow{ES}^i and \overrightarrow{ES}^f are the initial and final positions of the elbow-shoulder vector respectively, and \overrightarrow{EW}^i and \overrightarrow{EW}^f are the initial and final positions of the elbow-wrist vector respectively.


 figs/joint_diagram.png

Figure 13: Joint angle references. (W - wrist, E - Elbow, S - Shoulder, P - point in space below the shoulder joint, θ_E - elbow angle, and θ_S - shoulder angle)

When interacting with the *Super Pop VRTM* game, users are asked to sit in an upright position. This keeps their torso perpendicular to the transversal plane. Given that the game's virtual environment is tailored to the user's body dimensions (i.e. the virtual objects are at arm's reach), there is no need for users to move their torso to complete the reaching tasks. As such, any displacement in the position of the torso is negligible. This is to say that any displacement in the position of the shoulder-P vector (\overrightarrow{SP}), which is parallel to the user's torso, is also negligible - where the coordinates of the P point are defined by (??).

$$P_x = S_x + 0.5 (SO_x - S_x) \quad (16a)$$

$$P_y = S_y + 0.5 (SO_y - S_y) - 1 \quad (16b)$$

$$P_z = S_z + 0.5 (SO_z - S_z) \quad (16c)$$

where (S_x, S_y, S_z) and (SO_x, SO_y, SO_z) are the coordinates of the user's dominant and non-dominant shoulder respectively. Under this assumption, the shoulder ROM is a function of the initial and final positions of only the shoulder-elbow vector (\vec{SE}^i and \vec{SE}^f respectively). We consider any displacement that the \vec{SE} might have between the beginning and end of the reaching movement by translating both vectors to the origin. We then compute the angle between them which, by definition, is the shoulder ROM (??).

$$ROM_S = \cos^{-1} \left(\frac{\vec{SE}^i \cdot \vec{SE}^f}{\|\vec{SE}^i\| \|\vec{SE}^f\|} \right) \quad (17)$$

To determine how different the user's elbow and shoulder ROM values are from the model's ROM values, we compute a percent error between them normalized with respect to the maximum allowed angle displacement for the given joint (??).

$$ROM_{error} = \left| \frac{ROM_{user} - ROM_{model}}{ROM_{max}} \right| 100\% \quad (18)$$

where ROM_{error} is the percent error between the user's and model's ROM values, ROM_{user} and ROM_{model} are the effective ROM values for the user and model respectively, and ROM_{max} is the maximum allowed angle displacement for the given joint. We pragmatically defined these values for the elbow and shoulder joints to be 150° and 180° respectively.

4.3.3 Path Length

The Path Length (PL) parameter expands on the Deviation from Path parameter by providing more information as to how well an individual can control his/her movements. The PL of a trajectory is defined as the total length of its curve. An individual that has troubles with controlling their movements may move in a very different path than the calculated optimal one. As such, by comparing the length of the path taken by the user to the length of the path created by the model, the algorithm can better determine how well the user is able to control their movements. We compute a percent error between the user's and model's PL values to determine how far off the user's PL is from the model's (??). A large percent error means that the user took a longer path than needed to reach the target suggesting that the user is having some troubles with movement control. Similarly, a negative percent error suggests that the path taken by the user was shorter than the one created by the model, possibly because the user stopped moving at some points or because the movements were too slow to reach the target in time. In either case, the algorithm will determine the corresponding error and the therapist can provide the user with appropriate feedback.

$$PL_{error} = \frac{PL_{user} - PL_{baseline}}{PL_{baseline}} 100\% \quad (19)$$

4.3.4 Movement Time

A common symptom experienced by individuals who have any motor skills disorder is slow movements [?]. Therapists use the Movement Time (MT) parameter to analyze the user's speed. The MT for a given task in the Super Pop environment is the amount of time passed between the instances when the user 'pops' the START and TARGET bubbles. In general, for a given rehabilitation protocol, the idea is to compare the user's MT with a ground truth value to objectively determine how fast his/her movements are. Because of its wide adoption, we define a reference MT using

the model of human movement developed by Paul Fitts called **Fitt’s law** [?]. This model is used in human-computer interaction scenarios to predict the amount of time a user needs to move from one point to another in a virtual environment. Fitt’s law predicts MT as a function of the distance between the starting position and the target, and the width of the target. These two parameters are encompassed into a third parameter called the **difficulty index (DI)**, which measures the difficulty of a given task. The greater the distance and/or the smaller the target, the more difficult it becomes to complete the task. We adhere to the DI definition of the Shannon formulation because of its popularity [?], resulting in a model of the form (??).

$$DI = \log_2 \left(\frac{A_{2D}}{W} + 1 \right) \quad (20a)$$

$$MT = a + b DI \quad (20b)$$

where DI is the difficulty index [unitless], A_{2D} is the distance between the initial position and the target [pixels], W is the width of the target [pixels], MT is the predicted movement time [ms], and $a, b \in \mathbb{R}$ are the intercept and the slope of the final model respectively. The final time prediction model for a given task is as a linear function of its DI (??).

Constructing a Fitt’s model refers to training its slope and intercept to fit MT data collected from users interacting with the system, and it is highly dependent on the virtual environment (Figure ??). Thus, we constructed a 3D Fitt’s model specifically tailored for the *Super Pop VRTM* game [?]. The idea is to define a set of movement tasks, collect human MT data for all tasks, and correlate the MTs to the tasks’ corresponding DIs. Since we are interested in constructing a model that is appropriate for 3D movements, the distance traveled is now the 3D distance between the initial position of the user’s hand and the target. However, the movement tasks are defined based on the positions of the virtual bubbles which are defined in a 2D

pixel space. As such, we built two linear models. The first model correlates the 2D pixel distance between the virtual objects to the user's 3D path length (PL). The second model then correlates the DI of a task and the time needed to complete it. Thus, the final 3D Fitt's model remains a function of the 2D distance between the start and target points (A_{2D}), and the width of the target (W), while taking into consideration the users' 3D movements. The equation is given by (??), where a and b , and c and d are the intercepts and slopes of the MT and PL linear models respectively.

$$MT = a + b \log_2 \left(\frac{c + d A_{2D}}{W} + 1 \right) \quad (21)$$


 figs/FittsToSuperPop.png

Figure 14: Comparison between (a) reaching a target with a cursor in a computer environment, and (b) a reaching task the in the Super Pop environment. (Image (a) adapted from [?].)

We recruited seven able-bodied adults to interact with the *Super Pop VRTM* game. Sixteen tasks were empirically selected. Each participant was assigned to repeatedly complete eight randomly selected tasks. We collected, on average, **24 ± 5** PL and MT points for each task. The distances, widths, and DIs that define the tasks, and the collected PL and MT averages per task are shown in Table ??. To increase the correlation factor between variables for both models, we assume that both data sets follow a Gaussian distribution, and thus only considered data points that were within one standard deviation of the mean of the complete data set.

Table 2: MT and PL averages for each of the tasks used for constructing a 3D Fitt’s model.

Task	Distance [pixels]	Width [pixels]	DI [bits]	MT Avg [ms]	PL Avg [mm]
1	100.0	100	1.00	327.83	202.67
2	121.9	100	1.15	360.19	278.23
3	131.6	90	1.30	426.75	286.91
4	147.5	90	1.40	374.79	346.04
5	182.8	100	1.50	616.16	453.46
6	171.1	80	1.65	429.05	431.67
7	223.4	90	1.80	694.88	552.43
8	191.2	70	1.90	496.77	522.53
9	180.0	60	2.00	772.56	433.84
10	251.6	70	2.20	577.53	727.52
11	235.5	60	2.30	819.78	638.94
12	213.9	50	2.40	755.74	616.45
13	186.3	40	2.50	970.75	495.16
14	253.1	50	2.60	774.21	756.67
15	219.9	40	2.70	1181.74	606.11
16	238.6	40	2.80	1115.15	730.67

We performed a linear regression between the participants’ 3D PL and 2D pixel distance between the virtual objects (??) (Figure ??) ($R^2 = \mathbf{0.9703}$). We used this model to update the DI values for the defined tasks (Table ??), and construct the final Fitt’s time prediction model which is a correlation between the average MT values an their corresponding DIs (??) (Figure ??) ($R^2 = \mathbf{0.7299}$). The relatively high correlation values suggest that the models provide a good description of the observed behavior.

$$PL_{user} = -174.3 + 3.5651 A_{2D} \quad (22a)$$

$$MT_{user} = -25.044 + 346.68 \log_2 \left(\frac{PL_{user}}{W} + 1 \right) \quad (22b)$$

where A_{2D} is the 2D pixel distance between the start and target bubbles [unitless], PL_{user} is the user’s corresponding 3D PL [m], MT_{user} is the predicted movement time [ms], and W is the width of the target bubble [mm]. It’s important to note that the 3D PL model (??) is only used when constructing the Fitt’s model. During game play, the user’s baseline MT is computed as a function of his/her 3D PL directly, thus there is no need to convert the 2D pixel distance to a 3D PL value.

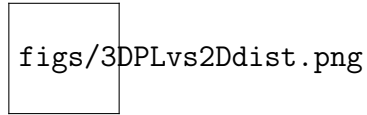


Figure 15: Three-dimensional PL averages of the collected human data versus 2D pixel distance between the start and target virtual objects, and the correlation between the 3D and 2D distances.

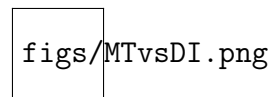


Figure 16: MT averages for the selected tasks, and the linear correlation between the averages and DIs.

Table 3: Updated DIs based on 3D PL model.

Task	Distance [pixels]	PL Model [mm]	Updated DI [bits]
1	100.0	182.2	0.79
2	121.9	260.3	1.03
3	131.6	294.9	1.21
4	147.5	351.6	1.36
5	182.8	477.6	1.54
6	171.1	435.6	1.67
7	223.4	622.1	1.92
8	191.2	507.5	1.97
9	180.0	467.4	2.05
10	251.6	722.8	2.36
11	235.5	665.2	2.45
12	213.9	588.3	2.52
13	186.3	489.8	2.56
14	253.1	728.2	2.78
15	219.9	609.7	2.83
16	238.6	676.2	2.96

To determine how different the user’s MT is from the model’s, we compute a ratio between both values (MT_{ratio}), where MT_{user} and MT_{model} are the user’s and model’s MTs respectively. Values of $MT_{ratio} > 1$ suggest that the user moves slower than the baseline by a factor of MT_{ratio} . Similarly, values of $0 < MT_{ratio} < 1$ suggest that the user moves faster than the baseline by a factor of MT_{ratio} . In this scenario, since the user completed the reaching task faster than the expected average, no error would be registered for that task.

$$MT_{ratio} = \frac{MT_{user}}{MT_{model}} \quad (23)$$

4.3.5 Movement Smoothness

Individuals who have some type of motor skills disorder tend to have jittery movements. Movement smoothness has been investigated as an indicator of motor skill and coordination [?]. Our proposed method measures how smooth/jittery the user’s movement are by computing the amount of movement units (MU) a given trajectory has [?, ?]. The lesser the amount of MUs, the smoother the movement. The amount of MUs a trajectory has is defined as a function of the curvature of the trajectory and the speed of the movement. Each unit consists of an acceleration and a deceleration

phase. Each new acceleration phase marks a new MU. Smooth linear movements yield one MU count (one acceleration and one deceleration phase). Given that the movement tasks in the Super Pop environment are generally completed with linear or slightly curved movements [?, ?], we define one MU as the baseline for the Movement Smoothness parameter. We don't compute a percent error for this parameter, but rather present the total amount of count a given trajectory has.

A new acceleration/deceleration phase is identified when the acceleration of the user's hand exceeds an upper threshold (TH) value AND goes below a lower TH value. It has been shown that the upper and lower TH values for identifying new acceleration/deceleration phases is highly dependent on the sensors used to measure the user's movements. In general, the TH values are usually pragmatically selected [?]. For our study, we defined an objective methodology that computes the TH values relative to the *Super Pop VRTM* platform. We collected acceleration profiles from 20 able-bodied adults. Fifteen males and five females ranging in age between 18 and 45 years (mean age = 28.4 years, standard deviation = 5.7 years) interacted with the Super Pop system. They were asked to complete the reaching task shown in Figure ?? three times with their dominant hand. Maintaining the assumption that linear movements are the baseline solution to the movement tasks ($MU = 1$), we updated the upper and lower TH values systematically such that all input acceleration profiles yielded a MU count of one. As such, any acceleration profile that does not fall within the computed TH values will yield a MU count greater than one. The steps of this script are described in Algorithm ??, where $getMU(Acc, UTH, LTH)$ is the function that computes the number of movement units in acceleration profile Acc , as a function of the upper and lower TH values, UTH and LTH respectively.

Algorithm 2 Compute the upper and lower thresholds for identifying new acceleration/deceleration phases.

Require: Acc_i : i^{th} ground truth acceleration profile ($\forall i = [1, \dots, N]$),
 UTH_o =initial upper threshold value, and LTH_o =initial lower threshold value

```

1:  $UTH \leftarrow UTH_o$ 
2:  $LTH \leftarrow LTH_o$ 
3: for  $i = 1 : N$  do
4:    $MU_c \leftarrow \text{getMU}(Acc_i, UTH, LTH)$ 
5:   while  $MU_c > 1$  do
6:      $d_u \leftarrow \|P_u - UTH\| \triangleright P_u$ : point in  $Acc_i$  greater than and closest to  $UTH$ 
7:      $d_l \leftarrow \|P_l - LTH\| \triangleright P_l$ : point in  $Acc_i$  smaller than and closest to  $LTH$ 
8:     if  $d_u < d_l$  then
9:        $UTH \leftarrow P_u$ 
10:    else if  $d_l < d_u$  then
11:       $LTH \leftarrow P_l$ 
12:    end if
13:     $MU_c \leftarrow \text{getMU}(Acc_i, UTH, LTH)$ 
14:  end while
15: end for
16: return  $UTH, LTH$ 

```

4.3.6 Average Speed

We expand on the Movement Time (MT) parameter by further analyzing the speed of the user's movements when completing a reaching task [?]. Although the Average Speed parameter provides similar information to the MT parameter, it presents it from a different perspective such that therapists can better assess the individual's kinematic performance. We determine how well users can control their movements by comparing their average movement speed to a speed value computed by our kinematic model. The baseline value, S_{model} , is computed by dividing the user's effected PL by the model's MT (??). The speed baseline is a function of the user's PL because the speed should be computed for the distance that the user traveled between the start and target points. This is to say, the system computes how fast the user's movement should be relative to the actual displacement generated. The user's average speed, S_{user} , is computed by dividing their total PL by their effective MT (??). We compute a

ratio to determine how much faster/slower the user completes a reaching task relative to the model's speed (??). Values of $S_{ratio} > 1$ suggest that the user moves faster than the baseline by a factor of S_{ratio} . In this scenario, since the user completed the reaching task faster than the expected average, no error would be registered for that task. Similarly, values of $0 < S_{ratio} < 1$ suggest that the user moves slower than the baseline by a factor of S_{ratio} .

$$S_{model} = \frac{PL_{user}}{MT_{model}} \quad (24a)$$

$$S_{user} = \frac{PL_{user}}{MT_{user}} \quad (24b)$$

$$S_{error} = \frac{S_{user}}{S_{model}} = \frac{MT_{model}}{MT_{user}} \quad (24c)$$

4.4 Pilot Study I: Range of Motion Comparison

In the spirit of validating the kinematic model and the definitions for the selected kinematic parameters, our first pilot study focuses on validating the kinematic model relative to the elbow and shoulder range of motion (ROM) parameters. The validation is made by analyzing the percent error differences between the participants' outcome metrics and the baseline values generated by the kinematic model. This section describes our first pilot study and presents updated results from our previous study described in [?].

4.4.1 Hypothesis

The following was the hypothesis for this study:

H_1 : Our proposed kinematic model (Section ??), accurately mimics the movement kinematics of the human arm relative to the elbow and shoulder ROM parameters (as defined in Section ??).

4.4.2 Experimental Design

Eleven typically developing children were recruited to interact with the *Super Pop VRTM* system. Six females and five males ranging in age between 6 and 11 years (mean age = 8.7 years, standard deviation = 1.7 years) played the game and were asked to follow the testing protocol described in Appendix ???. It's important to keep in mind that, although the movement task the participants were required to perform was the same for all, the state is correlated for each participant as defined by the corresponding body dimensions. The parents of the participants signed the IRB (Institutional Review Board) approved consent form allowing their children to engage in the testing sessions.

For each participant, a total of six reaching movements were collected per arm during game play. The 3D coordinates of the participants' upper-body joints are captured from the moment they 'pop' the START bubble until they 'pop' the TARGET bubble. The joint coordinates, target coordinates, and the participant's arm link lengths are used as input to: 1) compute his/her elbow and shoulder ROM nominal values (using the methods described in Section ??), and 2) generate the optimal trajectory that connects the START and TARGET bubbles (using the kinematic model described in Section ??). The baseline's ROM values are computed using the optimal trajectory's coordinates as input.

4.4.3 Results

The comparison between the participants' nominal ROM values versus the generated baseline values for a single trial, and the average errors of the measurements made are shown in Table ??. Taking into consideration the learning curve of the Super Pop platform, we show the resulting outcome metrics of the last trial of the participants' dominant arm. The progression of the elbow and shoulder ROM errors of Participant 1 over the six trials of their dominant arm is shown in Table ??. Boxplots showing

the variation of Participant 1's ROM percent errors are shown in Figure ??.

Table 4: Comparison between the participants' and model's values for a single trial.

Participants	Elbow ROM		Shoulder ROM	
	User [deg]	Error [%]	User [deg]	User [%]
1	27.45	10.74	46.27	17.59
2	27.65	12.45	34.16	12.20
3	7.38	4.42	31.58	2.46
4	6.62	2.10	25.84	2.12
5	27.38	17.88	20.09	9.15
6	0.23	4.38	19.31	3.18
7	16.93	3.01	36.28	1.22
8	-	-	-	-
9	2.92	2.63	21.73	0.99
10	3.27	1.63	17.11	2.68
11	5.06	1.71	47.63	2.93
AVG		6.10		5.45
STD		5.32		5.33

*Missing values are due to corrupt data in the collection process.

Table 5: Progression of elbow and shoulder ROM errors for Participant 1.

Trials	ROM Errors [%]	
	Elbow	Shoulder
1	14.67	17.91
2	3.83	6.75
3	-	-
4	1.05	1.87
5	12.60	12.42
6	10.74	17.59
AVG	8.58	11.31
STD	5.24	6.23

*Missing values are due to corrupt data in the collection process.

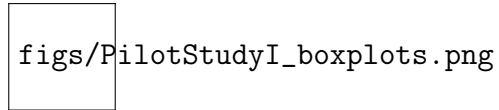


Figure 17: Boxplots showing the average elbow and shoulder ROM percent errors over six trials for Participant 1.

4.4.4 Discussion and Conclusions

From Table ??, participants yielded an average error of $6.10 \pm 5.32\%$ and $5.45 \pm 5.33\%$ for the elbow and shoulder joints respectively. From Table ?? and Figure ??, Participant 1 had an average error of $8.58 \pm 5.24\%$ and $11.31 \pm 6.23\%$ for the elbow and shoulder joints respectively. Information such as this is important to the clinician, which can be use to evaluate the kinematic performance of the population as a group or of the individuals independently and, if necessary, adapting the rehabilitation protocol to their needs.

Typical baseline models created by collecting human data as in [?, ?, ?], show an error ranging from 13.8% to 66.7%. Our results fall in-line with human-data collection approaches seeing as the average errors yielded by our approach are smaller than those in previous studies. Hypothesis H_1 is thus validated allowing us to conclude that our proposed kinematic model accurately mimics the movement kinematics of the human arm relative to the elbow and shoulder ROM parameters.

4.5 *Pilot Study II: Baseline Comparison*

Our second pilot study focuses on validating the baseline generated by our kinematic model. We compare the outcome measures generated by our model to the outcome measures obtained from a group of able-bodied adults. The comparison is made between all seven parameters described in Section ?. However, for this comparison we substitute the ‘Deviation from Path’ (DfP) parameter with the ‘**Deviation from Line**’ (DfL) parameter. Similar to the DfP parameter, the DfL parameter also computes the area between the two Euclidean trajectories to determine how different they are from each other (methods for computing the DfL cost value are the same as described in Section ?). The difference is that the DfL parameter evaluates a given trajectory by comparing it to the line segment that connects the start and end points, while the DfP parameter evaluates the trajectory by comparing it to the

model’s trajectory. We make this substitution because, when using the DfP parameter, the cost value of comparing the model’s trajectory with itself would always yield zero. As such, the DfL parameter provides unbiased information used to compare the two baselines.

4.5.1 Hypotheses

Not all the baseline outcome metrics are generated based on our kinematic model. The ground truth values for the deviation from line, path length, and elbow and shoulder range of motion parameters are relative to the kinematic model. The ground truth values for the movement time, movement smoothness, and average speed parameters are computed based on measurements from human models. For example, the baseline values for the movement time parameter are computed using our derived model based on Fitts law (Section ??). Given that the baseline outcome metric values are computed relative to different types of models, this study poses two hypotheses for the two different parameter groups:

H_1 : For the parameters based on the kinematic model, the differences between the average values generated by our kinematic model and the average values computed from human interactions with the *Super Pop VRTM* game are small enough such that we consider the two baselines similar to each other.

H_2 : For the parameters based on human models, the variability of the values generated from the respective models is lower than the variability of the values obtained from the participants that interacted with the game.

4.5.2 Experimental Design

Ten able-bodied adults were recruited to interact with the *Super Pop VRTM* system. Six females and four males ranging in age between 24 and 31 years old played the game and were asked to complete a 90° reaching task (Figure ?? - Appendix ??).

All participants signed the IRB (Institutional Review Board) approved consent form allowing them to participate in the testing sessions. In general, clinical studies focus on a specific participant demographic when constructing a baseline with human data. Thus, we took into consideration data from the subset of participants that have similar arm lengths to ensure that the movement exercises we evaluate are similar to each other. We selected the participants whose arm lengths fall within one standard deviation of the mean arm length from all original participants. Seven of the ten participants fell into this category.

Even after selecting the participants that have similar arm lengths, we observed that the results still had a relatively large amount of outliers due to the variability of human movement. It turns out that this is mainly due to the path length (PL) parameter. During game play, the system records that a bubble is ‘popped’ when the user’s hand reaches the bounds of the bubble. However, there are some cases where the camera does not have a high enough frame rate to keep track of the user’s movements (the Kinect camera’s frame rate is approximately $15fps$). This is to say that there are instances where the camera takes more than the necessary frames to capture the moment when the user reaches the target bubble even though he/she might have reached the bounds of the bubble in one of the previous frames. This results in a measured PL greater than what it should really be. To address this issue, we only considered the instances of tasks (from both the human and kinematic models) whose PL values were within one standard deviation of the PL population mean. This eliminates most of the outliers thus providing less biased data sets.

All participants were asked to complete a 90° reaching task (as described in Figure ??) ten times for each arm. We aggregated all of their resulting outcome measures and organized the data based on the corresponding kinematic parameter and on the arm that was used to complete the task. We did the same with the outcome measures computed by our kinematic model for each task. For each arm, this results in a vector

per parameter per model that we can compare with each other.

To investigate the relationship between the two generated baselines, for each parameter and for each arm, we computed the effect sizes (estimate difference between means) of the two models and their 99.99% confidence interval (CI) bounds (??). There is a probability of $1 - \alpha$ that the resulting interval will contain the true difference between the population means. By selecting $\alpha = 0.0001$ we're using the highest probability, thus computing the largest possible range of values that the resulting CI contains the true difference between the population means for each parameter. The following assumptions need to be met for (??) to be accurate: 1) the two populations have the same variance, 2) the populations are normally distributed, and 3) each instance is sampled independently from each other.

$$CI_{(1-\alpha)} = (\bar{X}_1 - \bar{X}_2) \pm (t_{\alpha/2}) (S_{\bar{X}_1 - \bar{X}_2}) \quad (25)$$

where \bar{X}_1 and \bar{X}_2 are the means of the vectors with instances from the human and kinematic models respectively, $t_{\alpha/2}$ is the t -distribution for $n_1 + n_2 - 2$ degrees of freedom, n_1 and n_2 are the lengths of the vectors for the human and kinematic models respectively, and $S_{\bar{X}_1 - \bar{X}_2}$ is the estimated standard error of the difference between the means. The true estimated standard error of the difference between means is given by (??). However, we don't know the true value of the population's variance (σ^2). Thus we estimate the standard error (MSE) by averaging the two sample variances (s_1^2 and s_2^2) (??). We then use this value to estimate the standard error defined as (??), where $n_1 = n_2 = n$, and $\sigma_1^2 = \sigma_2^2 = \sigma$ based on our original assumptions.

$$\sigma_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = \sqrt{\frac{2\sigma^2}{n}} \quad (26a)$$

$$MSE = \frac{s_1^2 + s_2^2}{2} \quad (26b)$$

$$S_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{2MSE}{n}} = \sqrt{\frac{s_1^2 + s_2^2}{n}} \quad (26c)$$

4.5.3 Results

Boxplots showing how the data are distributed for each parameter and for each arm are presented in Figure ???. The parameters left of the vertical line are relative to the kinematic model while the parameters right of the line are relative to additional human models. The sample means of the vectors with data obtained from the human and kinematic models are shown in Table ??, as well as the effect sizes \pm the window sizes of their 99% CIs.

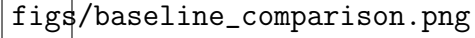
figs/baseline_comparison.png

Figure 18: Boxplots showing how the data are distributed between the collected human data and the data generated by our model for the right (top) and left (bottom) arms.

4.5.4 Discussion and Conclusions

For the parameters relative to the kinematic model, the effect sizes are $\sim 0^\circ$ for the EROM and SROM parameters for both arms (Table ??). The same is true for the DfL parameter. The effect sizes are $-0.004 \pm 0.010m^2$ and $-0.013 \pm 0.016m^2$ for the right and left arms respectively. Moreover, the 99.99% CI bounds $[\pm]$ are less than 5° for the ROM parameters, and in the range of $0.001 - 0.015m^2$ for the DfL parameter. The low bound values suggest that there is a 99.99% probability that the estimated effect sizes are the true mean differences between the two baselines relative to the selected parameters. Since effect sizes are ~ 0 , we conclude that there is a high probability

Table 6: Effect sizes between the baselines generated by collected human data and our kinematic model with their confidence intervals.

	Parameters	Means [Human Model]	Means [Kinematic Model]	Effect Sizes	99.99% CI Bounds [\pm]
Right Arm	DfL [$10^{-3}m^2$]	27.86	32.03	-4.16	9.62
	PL [mm]	346.84	289.83	57.01	42.63
	EROM [deg]	4.25	5.59	-0.34	2.36
	SROM [deg]	27.57	29.03	-1.46	4.02
	MT [ms]	574.61	385.86	188.65	122.86
	MUs [no units]	1.08	1.00	0.08	0.13
	Avg S [mm/s]	733.97	752.36	-18.39	186.87
Left Arm	DfL [$10^{-3}m^2$]	35.60	48.224	-12.63	15.62
	PL [mm]	398.18	309.76	88.43	59.59
	EROM [deg]	5.48	6.09	-0.61	2.90
	SROM [deg]	29.66	31.40	-1.74	4.90
	MT [ms]	556.28	423.46	141.82	144.93
	MUs [no units]	1.13	1.00	0.13	0.16
	Avg S [mm/s]	856.36	732.00	124.37	291.29

*DfL: deviation from line, PL: path length, EROM and SROM: elbow and shoulder range of motion, MT: movement time, MUs: movement units, Avg S: average movement speed

**Parameters above the dashed lines are relative to the kinematic model while parameters below the dashed lines are relative to additional human models.

that the two baselines are similar to each other with respect to the DfL, EROM, and SROM parameters. Regarding the PL parameter, the effect sizes are $57.01 \pm 42.63mm$ and $88.43 \pm 59.59mm$ for the right and left arms respectively. Although considerably small effect sizes, the CI bounds are relatively large. This indicates that the true mean difference between the two baselines for the PL parameter could be greater than $100mm$ for either arm. As such, our first hypothesis (H_1) is supported for the DfL, EROM, and SROM parameters but not necessarily for the PL parameter. This is to say that the differences between the sample means of the data collected from humans and the data computed by our kinematic model are small enough such that we consider the two baselines statistically similar to each other with respect to the DfL, EROM, and SROM parameters.

Regarding the parameters relative to the external human models, we did not expect the baseline generated by our models to be equivalent to the baseline created with human data. This is because the external human models were not designed to generate results by mimicking the human movement as we did for the previous parameters. Instead of focusing on the effect sizes and their CIs, we rather look at the variability

of the data. The variability of the data collected from the participants is greater than the variability of the data computed by the corresponding models for each parameter (Figure ??). This is especially true for the MT and AvgS parameters. This is to say that, a baseline constructed with data computed from external human models would yield a more accurate description of how humans complete movement tasks when compared to a baseline constructed from the data directly collected from participants because of its high variability. For example, instead of aggregating movement times from different participant demographics, our Fitts model predicts the amount of time needed to complete a movement task as a function of the difficulty of the task. Although the Fitts model was also constructed with human data, it reduces the participants' natural variability by masking it in a linear function. As such, our second hypothesis (H_2) is met for the MT, MU, and AvgS parameters.

The validation of the two hypotheses allow us to thus validate the baseline generated by our kinematic model. Namely because: 1) it yields similar values as a model constructed with human data with respect to the DfP, EROM, SROM, and PL parameters, and 2) it yields ground truth values with less variability than the one generated by the baseline constructed with human data with respect to the MT, MU, and AvgS parameters.

4.6 Pilot Study III: Super Pop VRTM as an Evaluation Tool

Previous studies have shown that using reaching kinematics measured by motion analysis yields larger quantifiable improvements than studies which used standardized clinical assessment tools when comparing pre- and post-interventions [?]. However, given that current state-of-the-art motion analysis systems (e.g. Vicon and OptiTrack systems) are expensive and are typically located in the clinical setting due to their lack of portability, we developed the *Super Pop VRTM* game; a tool that not only capitalizes on the benefits of using reaching kinematics measured by motion analysis,

but also increases the benefits by allowing for use in the in-home setting (refer to Chapter ?? and Chapter ?? for the benefits of in-home rehabilitation protocols and the description of the serious game respectively). When developing a new tool, it's important to evaluate its efficacy and efficiency. We conducted a separate clinical study [?] to determine if the *Super Pop VRTM* game is efficient for documenting arm function improvement in children who have cerebral palsy relative to their reaching kinematics, and to verify the results from [?] that support the use of reaching kinematics measured by motion analysis as the basis for assessing patients' performances through their intervention protocols.

4.6.1 Hypotheses

The following were the hypotheses for this study:

H_1 : The *Super Pop VRTM* game is a feasible system for the in-home setting for documenting arm function improvement in children who have cerebral palsy relative to their reaching kinematics.

H_2 : Analysis of the reaching kinematics measured by the *Super Pop VRTM* game correlate with the analyses obtained from standardized clinical assessment tools.

4.6.2 Experimental Design

Three children with cerebral palsy (CP) (three girls, mean age = 9 years, standard deviation = 1.73 years), and 11 typically developing children (six girls and five boys, mean age = 8.87 years, standard deviation = 1.87 years) were recruited to participate in this study. The parents of all participants signed the IRB (Institutional Review Board) approved consent form allowing their children to engage in the testing sessions. The group of children with typical development played the *Super Pop VRTM* game once (Appendix ??) and their outcome metrics served as the 'norm' comparison. The group of children with CP received an 8-week virtual reality (VR) intervention

using the commercial EyeToy Play VR system. Their parents were asked to record their child’s playing time for each session. In addition, the children were also required to maintain their regular physical and occupational therapy sessions throughout the intervention period. They were evaluated three times: before the intervention, mid intervention (after 4 weeks of the intervention), and immediately after the intervention. Three types of measurements were used: 1) reaching kinematics using the *Super Pop VRTM* game, 2) the fine motor scale of the Bruininks-Oseretsky Test of Motor Proficiency second edition (BOT-2) [?], and 3) the Pediatric Motor Activity Log (PMAL) [?].

Regarding the reaching kinematics, we focus on summarizing the results obtained from *t*-test analyses. For details on additional results refer to [?]. When measuring the reaching kinematics, participants were asked to follow the protocol described in Appendix ???. For this study we computed the following kinematic parameters: elbow and shoulder range of motion (Section ??), path length (Section ??), movement time (Section ??), number of movement units (Section ??), and the average hand speed (Section ??). All *t*-tests were performed at a 90% confidence level ($\alpha = 0.10$) to compare the mean values of reaching kinematics in children with CP and those in children with typical development.

Regarding the standardized clinical assessment tools, the BOT-2 assesses proficiency in fine manual control, manual coordination, body coordination, and strength and agility composite. For this study, the total point scores on fine motor precision, fine motor integration, and manual dexterity were computed and converted to *z*-scores via a scale score as suggested by the BOT-2 manual [?]. If the absolute change in *z*-score between pre- and post-test is greater than 1.65, the change is considered to be statistically significant at the 90% confidence level; which is equivalent to a *p*-value being less than 10% ($p < 0.10$). The PMAL allows parents to provide a rating about their child’s affected arm on 22 arm-hand real-world functional activities (e.g. holding

a cup, taking off shows, and turning a knob). Parents indicate “how often” their child used the more affected hand for each activity, and “how well” their child completed these functional activities. The PMAL score of each child with CP was compared with the minimal detectable change (MDC) reported from the literature. The scores greater than the corresponding MDC are considered to be clinically meaningful [?]. More information on the BOT-2 and PMAL assessment methodologies can be found in Section ??.

4.6.3 Results

The reaching kinematics of the more affected hand in children with and without CP are shown in Table ?. In comparing the mean values of the reaching kinematics in children with CP and those in typically developing children, the independent t -tests yielded the following results; At **pre-test**, children with CP had longer path length ($p = 0.005$), longer movement time ($p = 0.002$), more number of movement units ($p = 0.01$), and smaller shoulder range of motion ($p = 0.01$) than children with typical development. There were no differences on elbow range of motion and average speed of the hand ($p > 0.10$). At **mid-test**, children with CP still had longer MT ($p = 0.01$) and more number of movement units ($p = 0.01$) than children with typical development, but showed no differences on any of the other kinematic variables ($p > 0.10$). At **post-test**, none of the variables showed statistically significant differences ($p > 0.10$). These results are summarized in Table ?. The BOT-2 z -scores and PMAL scores obtained from children with CP are shown in Table ? and Table ? respectively.

Table 7: Outcome measures of the more affected hand in children with cerebral palsy (CP) and in typically developing (TD) children (adapted from [?]).

Participants		PL [m]	MT [s]	MUs [no units]	AvgS [m/s]	EROM [deg]	SROM [deg]
C1	Pre-test	0.95	2.41	6.50	0.38	21.53	51.75
	Mid-test	0.55	1.22	3.78	0.44	15.34	29.94
	Post-test	0.63	1.37	4.67	1.02	29.65	23.63
C2	Pre-test	0.44	0.95	2.71	0.72	10.07	36.02
	Mid-test	0.50	1.27	3.75	0.38	15.05	28.79
	Post-test	0.30	0.65	1.88	1.07	9.28	23.55
C3	Pre-test	1.41	3.65	5.29	0.46	32.50	58.43
	Mid-test	0.51	1.02	5.15	0.57	24.31	52.68
	Post-test	0.33	0.88	1.00	0.37	14.87	25.75
TD Children [AVG]		0.43	0.80	2.23	0.61	16.25	35.49
TD Children [STD]		0.17	0.26	1.06	0.24	8.88	9.79
TD Lower limit		0.097	0.290	0.152	0.140	-1.155	16.302
TD Upper limit		0.763	1.310	4.308	1.080	33.655	54.678

*PL: path length, MT: movement time, MUs: movement units, AvgS: average speed, EROM and SROM: elbow and shoulder range of motion.

Table 8: Summary of the results showing which kinematic parameters hold a statistical difference between the mean values of the kinematics in children with CP and those with typical development, between the pre-, mid-, and post-test.

	Kinematic Parameters					
	PL	MT	MUs	AvgS	EROM	SROM
Pre-test	✓	✓	✓	✗	✗	✓
Mid-test	✗	✓	✓	✗	✗	✗
Post-test	✗	✗	✗	✗	✗	✗

*PL: path length, MT: movement time, MUs: movement units, AvgS: average speed, EROM and SROM: elbow and shoulder range of motion.

✓: there is a statistical difference between the group of children with CP and without

✗: there is no statistically significant difference

4.6.4 Discussion and Conclusions

Results of the case series validated our first hypothesis (H_1) by demonstrating the feasibility of using the *Super Pop VRTM* system as a home-based evaluation tool for kinematic metrics in children with cerebral palsy (CP). It was successfully used with all three children with CP to collect the desired reaching kinematics in their natural environment (Table ??). The children enjoyed playing the *Super Pop VRTM* game without noticing that their reaching movements were recorded and quantitatively measured. They were quite cooperative during evaluation as they considered the whole procedure as part of the computer game playing experience. In general, the

Table 9: The scores on BOT-2 in children with CP (adapted from [?]).

ID	Fine Motor Precision			Fine Motor Integration			Dexterity		
	Pre	Mid	Post	Pre	Mid	Post	Pre	Mid	Post
C1	-2.4	-2.6	-2.6	-2.4	-2.4	-2.6	-2.8	-2.6	-2.4
C2	-2.8	-2.6	-2.8	-2.4	-2.4	-2.4	-2.6	-2.4	-2.6
C3	-2.4	-2.8	-2.6	-2.0	-2.8	-2.4	-2.6	-2.8	-1.6

Table 10: The scores on PMAL in children with CP as logged by the children’s parents (adapted from [?]).

ID	How Often			How Well		
	Pre	Mid	Post	Pre	Mid	Post
C1	4.636	5.000	5.000	4.591	4.955	4.955
C2	4.773	4.909	4.909	4.682	4.818	4.818
C3	2.045	2.500	2.955*	1.955	2.045	3.182*

The change score needs to exceed 0.67 for “how often”, and 0.66 for “how well” to be clinically meaningful.

* indicates that the PMAL score exceeded the minimal detectable change between the pre- and post-intervention evaluation.

system was easy to assemble and implement. On average, it took less than 10 minutes to set up in various home environments. The complete testing duration for the testing protocol (Appendix ??) took around 20 minutes, which gave us nine reaches per hand per participant, enough for an accurate evaluation of their kinematic performance throughout the intervention. After receiving the intervention, all children with CP improved their reaching kinematics between the pre- and post-test evaluations as measured by comparing their kinematic performance with typically developing children who provided the “norm” data (Table ??). Namely, at the post-test evaluation there was no statistically significant difference between mean values of the kinematic parameters in children with CP and those with typical development (Table ??). For more detailed results regarding the improvement of the children’s arm function using the z score analysis on the reaching kinematics refer to [?].

Our second hypothesis (H_2) could not be validated seeing as, when using the standardized assessment tools to measure improvement, only child *C3* improved her manual dexterity of BOT-2, although none showed any statistical improvement, (Table ??), and had parents rating an improvement on “how well” and “how often” she

used her affected hand in daily living. More specifically, all three children with CP increased their PMAL scores in both categories (Table ??). However, only child C3 had a change in score between pre- and post-intervention evaluation larger than the minimal detectable change reported in the literature. Thus, the improvements observed relative to the measured reaching kinematics do not correlate with the improvements relative to the clinical assessment methodologies. Several possible explanations are proposed in [?]; (1) The virtual reality intervention used in this study had more emphasis on repeated practicing reaching movements in all directions. However, this kind of “specificity of training” in reaching movements might not be able to generalize to other hand-arm functional activities, which were measured in BOT-2 and PMAL. (2) Other potential reasons for the disparity between the assessment methodologies are the insensitivity of the BOT-2 fine motor domain and/or the subjectivity of the BOT-2 and PMAL tools. Both methodologies can have troubles detecting minor changes seeing as the resulting scores depend on the observations of quality improvement made by the therapist or parent.

The study had a few limitations. First, the numbers of children with CP and children with typical development were small with wide variability in the children with CP. Future studies should increase the number of children with and without CP. Second, although inexpensive commercial VR gaming systems increase the accessibility of VR for training children with CP, there were challenges in applying VR systems designed for recreation to do rehabilitation. For example, the level of difficulty of the games and the content of the games (e.g. some violent scenes or age-inappropriate language) might not be suitable for some players. Therefore, future studies may need to evaluate the effects of a tailored intervention using our *Super Pop VRTM* game. Third, our evaluations were limited to pre-, mid-, and post-test. Children’s performance might be variable for a number of reasons, such as fatigue and sickness. A better home-based evaluation system should have the ability to evaluate the child’s

performance on a daily basis. Our *Super Pop VRTM* game has the potential to serve as a tailored VR intervention as well as a daily evaluation system. Therefore, in future research studies we will use the *Super Pop VRTM* game to conduct a tailored VR intervention program for children with CP and examine their reaching performance on a daily basis to examine longitudinal improvement in reaching movements in children with CP.

4.7 Summary

This chapter discussed the importance of comparing an individual’s movement characteristics against a baseline not only to determine the effectiveness of an intervention protocol, but also to objectively and quantifiably assess the individual’s progress in real-time. In the spirit of developing an in-home non-contact based rehabilitation system, we developed an assessment methodology to be coupled with the *Super Pop VRTM* game (Chapter ??), allowing for the evaluation of the user’s kinematic performance during game play. For each completed movement task, the system does so by comparing the user’s outcome metrics to the corresponding ground truth values. Instead of physically collecting human data for constructing said baseline, which can be time consuming and difficult to repeat for all patient demographics or all movement characteristics, we modeled the kinematics of the human arm (Section ??). For a given task, the final model generates the optimal solution in real-time as a function of the user’s arm dimensions and the state of the task. This chapter also defined and described a set of kinematic parameters that mathematically describe an individual’s upper-body movements (Section ??). These are used as the basis for assessing the user’s kinematic performance. A summary of the parameters’ symbols and general definitions are provided in Table ??.

This chapter concluded with three separate pilot studies to validate the kinematic model and parameters, and the feasibility of the overall Super Pop system.

Table 11: Summary of the kinematic parameters and their definitions.

Name	Symbol	Definition
Deviation from Path	DfP	Determines how far off the user’s movements are, spatially, from the optimal trajectory of the movement task.
Elbow Range of Motion	EROM	Elbow flexibility. Determines the extent of the range of movement an individual has for his/her elbow.
Shoulder Range of Motion	SROM	Shoulder flexibility. Determines the extent of the range of movement an individual has for his/her shoulder.
Path Length	PL	The PL of a trajectory is the total length of its curve. It measures how well an individual can control his/her movements.
Movement Time	MT	The total amount of time taken to move between the start and target virtual bubbles.
Movement Smoothness	MUs	Measures how jittery an individual’s movements are. It is a measurement of motor skill and coordination.
Average Movement Speed	AvgS	The average speed when moving between the start and target virtual bubbles.

The first study focused on validating the kinematic model relative to the elbow and shoulder (ROM) parameters (Section ??). We analyzed the percent error differences between the participants’ outcome metrics and the baseline values generated by the kinematic model. Results showed that participants yielded an average error of **6.10 \pm 5.32%** and **5.45 \pm 5.33%** for the elbow and shoulder joints respectively. The study concluded that, because the results fall in-line with error ranges obtained from human-data collection approaches (e.g. 13.8% to 66.7% [?]), the study’s hypothesis was validated allowing us to make the claim that the proposed kinematic model can accurately represent the movement kinematics of the human arm relative to the elbow and shoulder ROM parameters.

The purpose of the second pilot study was to validate the baselines generated by our system with respect to all seven kinematic parameters defined in this chapter (Section ??). For the kinematic parameters based on our kinematic model, results showed that the differences between the sample means of the data collected from humans and the data computed by our kinematic model are small enough such that we consider the two baselines statistically similar to each other with respect to the

deviation from line, and elbow and shoulder range of motion parameters, thus validating the study’s first hypothesis. For the kinematic parameters that are based on human models, results showed that, a baseline constructed with data computed from external human models would yield a more accurate description of how humans complete movement tasks when compared to a baseline constructed from the data directly collected from participants because of its high variability, thus validating the study’s second hypothesis.

Finally, the third pilot study focused on: 1) validating the *Super Pop VRTM* game as a feasible system for the in-home setting for documenting arm function improvement in children who have cerebral palsy (CP) relative to their reaching kinematics, and 2) validating the analysis reaching kinematics measured by the *Super Pop VRTM* game with respect to two standardized clinical assessment methodologies (Section ??). One of the two hypotheses were validated. The study validated the first hypothesis concluding that the *Super Pop VRTM* game is feasible for documenting improvement in children’s reaching kinematics seeing as the system reliably kept track of the participants’ reaching kinematics (Table ??), the system was easy to assemble and implement, and the participants enjoyed playing the game without noticing that their movements were quantitatively measured. Moreover, results showed that there was no statistically significant difference between mean values of the kinematic parameters in children with CP and those with typical development, after an 8-week home-based VR intervention for improving children’s arm function (Table ??). The second hypothesis could not be validated seeing as the observations made by analyzing the participants’ reaching kinematics do not match the observation made by the BOT-2 and PMAL assessment methodologies. While the analysis on the reaching kinematics suggests that all participants improved their arm-function, the BOT-2 and PMAL tools suggest that only participant *C3* improved her manual dexterity on BOT-2, although not statistically significant, (Table ??), and had parents rating

an improvement on “how well” and “how often” she used her affected hand in daily living, as measured by PMAL (Table ??). Potential reasons that explain the disparity include: 1) the VR training evaluated in this study might not be able to generalize to other hand-arm functional activities, which were measured by BOT-2 and PMAL, and 2) the insensitivity and/or subjectivity of BOT-2 and PMAL.

CHAPTER V

VALIDATION OF SUPER POP VRTM ACCURACY AND RELIABILITY

5.1 *Introduction*

Previous chapters describe the novelty and functionality of our *Super Pop VRTM* game (Chapter ??), and validate the feasibility of coupling the game with an assessment methodology to use it as an evaluation tool in intervention protocols (Chapter ??). As previously described, the game uses the KinectTM camera from Microsoft as its motion analysis system such that it can be used in the home environment. As such, this chapter focuses on validating the game's accuracy and reliability relative to its sensing methods. Section ?? compares the performance of the game and assessment methodology while using the Kinect versus using a current state-of-the-art motion capture system. We support our claim that the Kinect can be used as a sensor input to accurately evaluate users' reaching kinematics as measured by the *Super Pop VRTM* game by showing that it can yield similar performance as the highly accurate motion capture system relative to computed kinematic outcome measures. Moreover, when developing a new evaluation tool, it's important to examine its test-retest reliability within and between days. Section ?? examines the game's test-retest reliability with respect to the measurement reaching kinematics in able-bodied adults.

5.2 *Pilot Study: Accuracy of Sensing Methods*

Physiotherapists have been combining virtual reality (VR) systems with motion capture systems to keep track of users' movements. Currently, the most precise technology are marker-based systems, like the Vicon and OptiTrack systems. These use

optical sensors to track reflective markers attached to the user’s body and determine the three-dimensional (3D) coordinates of the markers, thus being able to locate the position of the user’s body and limbs. Although shown to be highly accurate, these marker-based systems are limited when applied to the home setting. Some limitations include: cost, the necessity for a controlled environment, and the time required for marker placement. As such, marker-based motion capture systems have generally been limited to the clinical setting.

This is one of the main reasons why the KinectTM camera from Microsoft, a portable 3D motion capture system, has been adopted in a rising number of VR physiotherapy games. In this section we evaluate the *Super Pop VRTM* game by comparing its performance while using the Kinect versus while using a marker-based motion capture system: the OptiTrack. Previous studies have compared and validated the Kinect’s accuracy relative to marker-based motion capture systems. For instance, C. Chang et al. [?] showed that the Kinect can achieve competitive motion tracking performance similar to the OptiTrack. However, the study focused only on spatial comparisons. C. Metcalf et al. [?] showed that the Kinect is suitable for home-based motion capture by comparing its performance versus the Vicon system, but focused only on assessment of the hand and finger joint kinematics. R. Clark [?] showed that the Kinect and the Vicon systems have comparable inter-trial reliability and excellent concurrent validity for the majority of measurements, but focused only on assessment of postural control. In this section we further investigate the Kinect’s capabilities as a sensor for tracking the user’s upper-body movements, and measuring and evaluating the set of kinematic parameters previously defined in Section ?? . Further details of the study can be found in [?, ?].

5.2.1 Hypotheses

We evaluate our system’s ability by comparing its performance while using the Kinect versus while using a state-of-the-art maker-based motion capture system: the OptiTrack system. In this manner, we can support our claim that the Kinect camera can be used as a motion capture system for home-based rehabilitation as well as identify any limitations with its use. The hypotheses of this study are:

H_1 : The differences between the outcome metrics computed with data collected with the Kinect camera and the outcome metrics computed with data collected with the OptiTrack system are sufficiently small such that we consider the results from the two systems to be equal to each other.

H_2 : The trajectories generated with Kinect and OptiTrack data are spatially similar to each other with respect to the hand, elbow, and shoulder joints.

5.2.2 Experimental Design

Ten able-bodied adults were recruited to interact with the *Super Pop VRTM* system. Six females and four males ranging in age between 24 and 31 years old played the game and were asked to complete a 90° reaching task (as described in Figure ??) ten times for each arm (Appendix ??). All participants signed the IRB (Institutional Review Board) approved consent form allowing them to participate in the testing sessions. Their interactions were recorded with both the Kinect camera and the OptiTrack system. The participants were asked to wear a non-infrared reflective suit with passive infrared markers attached to it for the extraction of the OptiTrack data. The reflective suit had 37 markers in total (Figure ??). Our OptiTrack setup consists of six *Flex 3* cameras. The layout of the testing environment is shown in Figure ?. Details about how to calibrate the OptiTrack cameras can be found in [?].

To validate the first hypothesis (H_1), we focus on the kinematic parameters that depend only on the tracking system: path length (PL) (Section ??), and elbow and

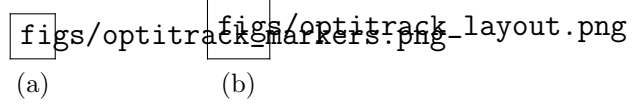


Figure 19: (a) IR marker configuration, and (b) layout showing placement of the user and the Kinect and OptiTrack cameras.

shoulder range of motion (EROM and SROM respectively) (Section ??). We do not consider the parameters that depend on additional external variables because these can potentially introduce errors that are not derived from the two systems directly. For example, the movement time parameter (Section ??) is independent of the tracking systems' capabilities. For each participant, we computed a percent error difference between the outcome metrics computed with Kinect and OptiTrack data. We computed a percent error value for each movement task the participant completed and averaged the values per arm. The final percent error difference per participant is computed using (??).

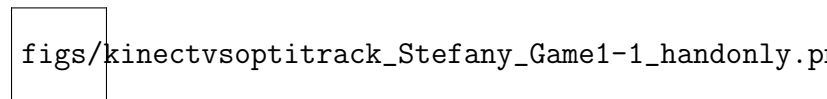
$$PE^{p,\alpha,\beta} = \frac{1}{n} \sum_{t=1}^n \frac{|V_K^{p,\alpha,\beta}(t) - V_O^{p,\alpha,\beta}(t)|}{nf^\beta} 100\% \quad (27)$$

where $PE^{p,\alpha,\beta}$ is participant p 's average percent error difference of parameter β for n completed movements tasks with arm α , $V_K^{p,\alpha,\beta}(t)$ and $V_O^{p,\alpha,\beta}(t)$ are the outcome measures computed with Kinect and OptiTrack data respectively for participant p , parameter β , and trial t , and nf^β is the normalization factor for parameter β . Each parameter has its own normalization factor. The EROM and SROM parameters are normalized with respect to their maximum allowed ROM (150° and 180° respectively [?]). Given that there is no maximum value allowed for the PL parameter (i.e. any trajectory can have an infinite length in theory), its percent error is normalized with respect to the value computed with OptiTrack data since we consider it to be the ground truth value.

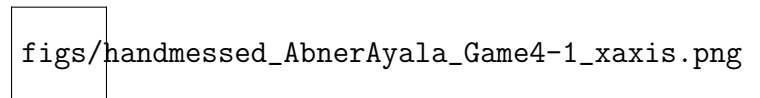
To validate the second hypothesis (H_2), we use the deviation from path (DfP)

parameter (Section ??) to quantify the similarity between the trajectories generated with Kinect and OptiTrack data for the hand, elbow, and shoulder joints. For each participant, each arm, and each joint, we compute and average the area between the Kinect and OptiTrack curves for each completed movement task.

Before performing the computations, we first eliminate the trials with corrupt data. For the OptiTrack, corrupt data occurs when the cameras lose track of one or more of the suit's IR markers. For the Kinect, this happens when the camera loses track of one or more of the user's joints. When there is an incorrect estimate of the user's position, inaccurate joint coordinates are stored. Examples of trials without and with corrupt data are shown in Figure ??.



(a) Trajectories without corrupt data.



(b) Hand trajectory with corrupt OptiTrack data.

Figure 20: Examples of Kinect (blue) and OptiTrack (red) trajectories without (a), and with (b) corrupt data.

5.2.3 Results

The average percent error differences between the outcome measures computed with Kinect data and the outcome measures computed with OptiTrack data per participant

for the PL, EROM, and SROM parameters are shown in Table ???. The average area values between the trajectories generated with Kinect and OptiTrack data per participant for the hand, elbow, and shoulder joints are shown in Table ???.

Table 12: Average percent errors for each parameter per participant.

Participants	Right Arm PE [%]			Left Arm PE [%]		
	PL	EROM	SROM	PL	EROM	SROM
1	15.1	1.9	2.0	12.9	7.1	0.8
2	-	-	-	-	-	-
3	8.3	3.3	3.0	10.7	3.2	4.4
4	19.4	5.0	9.0	34.5	5.1	9.5
5	12.3	3.7	2.2	10.2	6.5	1.0
6	4.3	2.8	3.4	3.7	2.2	1.4
7	6.6	9.7	7.6	-	-	-
8	5.2	13.1	20.1	5.9	5.9	7.6
9	4.2	8.8	13.3	6.3	14.4	11.8
10	19.0	6.1	3.0	15.7	5.9	7.5
AVG	10.5	6.0	7.1	12.5	6.3	5.5
STD	6.1	3.8	6.2	9.7	3.7	4.2

*PE: percent errors, PL: path length, EROM and SROM: elbow and shoulder range of motion.

**Missing values are due to all trials having corrupt data.

Table 13: Average areas as computed by the DfP parameter for the hand, elbow, and shoulder joints per participant.

Participants	Right Arm DfP [$10^{-3} m^2$]			Left Arm DfP [$10^{-3} m^2$]		
	Hand	Elbow	Shoulder	Hand	Elbow	Shoulder
1	62.2	48.4	38.9	48.1	46.8	19.8
2	-	-	-	-	-	-
3	68.2	34.3	41.2	79.2	47.1	39.7
4	111.4	56.1	36.0	167.8	62.9	25.2
5	95.1	57.3	42.4	132.5	62.2	52.1
6	80.5	53.7	53.1	76.5	49.1	41.7
7	58.6	94.5	45.3	-	-	-
8	108.4	38.5	37.9	116.1	50.9	34.3
9	102.8	91.5	93.6	73.7	88.0	63.7
10	116.1	50.1	37.0	175.9	120.2	41.6
AVG	89.3	58.3	47.3	108.7	65.9	39.8
STD	22.3	21.1	18.1	47.0	25.9	14.0

*DfP: deviation from path

**Missing values are due to all trials having corrupt data.

5.2.4 Discussion

The percent errors averaged from all participants for the PL, EROM, and SROM parameters are relatively low (Table ???). Results support our first hypothesis (H_1), suggesting that the Kinect is a viable option for home-based rehabilitation because it provides similar accuracy as the highly accurate current state-of-the-art marker-based

motion capture system within **5-7%** for the EROM and SROM parameters, and **10.5-12.5%** for the PL parameter. Similarly, the areas averaged from all participants for the hand, elbow, and shoulder joints are in the order of less than $0.15m^2$ (Table ??). These low values support our second hypothesis (H_2), that the trajectories generated by both tracking systems are spatially similar to each other. This further supports the validity and feasibility of using the Kinect camera as the motion capture system for home-based rehabilitation purposes.

5.3 Test-Retest Reliability in the Super Pop VRTM Game

In Section ?? we evaluated the *Super Pop VRTM* game’s efficacy by showing that it is a feasible tool for documenting arm function improvement in children who have cerebral palsy relative to their reaching kinematics. This section focuses on evaluating the game’s reliability and consistency of repeated responses. Previous studies have shown that, when developing a new tool, it’s also important to examine its test-retest reliability and consistency of repeated responses over time [?, ?]. As such, we conducted a study with the main goal of determining the test-retest reliability of reaching kinematics measured by the *Super Pop VRTM* game on a single day and on different days in a group of able-bodied adults. For more details on the full studies with adults, typically developing children, and children who have cerebral palsy refer to [?, ?, ?]. In addition, these studies also focus on determining the response stability and the minimum detectable change of reaching kinematics as measured by the game.

5.3.1 Experimental Setup

Twenty able-bodied young adults were recruited to interact with the *Super Pop VRTM* game. Seven females and 13 males ranging in age between 20 and 40 years (mean age = 26.55 years, standard deviation = 3.12 years) played the game and were asked to complete a 90° reaching task (as described in Figure ??) ten times for each arm (Appendix ??). All participants signed the IRB (Institutional Review Board) approved

consent form allowing them to participate in the testing sessions. Each participant underwent two assessments separated by a range of two to five days (3.30 ± 0.86 days). On each assessment day, participants played three games with ten reaching tasks for each arm. For each reaching task, the assessment algorithm described in Chapter ?? computed the following kinematic parameters: elbow and shoulder range of motion (EROM and SROM respectively) (Section ??), path length (PL) (Section ??), movement time (MT) (Section ??), number of movement units in the trajectory (MUs) (Section ??), and average hand speed (AvgS) (Section ??). All statistical tests were performed using two-tailed tests with a 95% confidence level.

We calculated the test-retest reliability between two trials for each testing day and between the two testing days for both arms using the **intraclass correlation coefficient (ICC)**, Model 2 [?, ?]. In general, the ICC is a measure of the reliability and consistency of measurements [?]. Model 2 assumes each participant was assessed by the same group of raters. For this study, the reaching kinematics as measured by the *Super Pop VRTM* game serve as the participants' raters. The ICC is a score between 0 and 1, where 0 indicates no reliability and 1 indicates perfect reliability. For this study, ICC values were interpreted as poor (< 0.40), fair to good ($0.40 - 0.74$), or excellent (> 0.75) [?, ?].

5.3.2 Results

The outcome measures averaged from both arms, and their corresponding ICC values within days and between the two testing days are summarized in Table ?? and Table ?? respectively.

5.3.3 Discussion and Conclusions

Overall, the *Super Pop VRTM* game is reliable for evaluating its users' reaching kinematics within the same day and between days. The test-retest reliability between games within each testing day was excellent for most parameters (from **0.783** to

Table 14: Test-retest reliability of reaching kinematics average *within* the same testing days.

	Parameters	Trial 1	Trial 2	ICC [95% CI]
Day 1	Path Length [m]	0.372 \pm 0.054	0.363 \pm 0.054	0.935 [0.836, 0.974]
	Movement Time [s]	0.619 \pm 0.213	0.524 \pm 0.218	0.894 [0.451, 0.967]
	Movement Units [no units]	1.106 \pm 0.096	1.124 \pm 0.119	0.487 [-0.322, 0.798]
	Average Speed [m/s]	0.713 \pm 0.248	0.893 \pm 0.354	0.793 [0.222, 0.930]
	Elbow ROM [deg]	4.182 \pm 1.867	3.836 \pm 1.274	0.713 [0.286, 0.885]
	Shoulder ROM [deg]	31.514 \pm 4.275	29.946 \pm 3.812	0.866 [0.576, 0.951]
Day 2	Path Length [m]	0.361 \pm 0.058	0.349 \pm 0.065	0.866 [0.576, 0.0951]
	Movement Time [s]	0.537 \pm 0.254	0.500 \pm 0.248	0.963 [0.902, 0.986]
	Movement Units [no units]	1.135 \pm 0.135	1.123 \pm 0.118	0.371 [-0.659, 0.755]
	Average Speed [m/s]	0.876 \pm 0.342	0.923 \pm 0.426	0.938 [0.846, 0.975]
	Elbow ROM [deg]	4.688 \pm 2.511	4.715 \pm 2.992	0.958 [0.846, 0.975]
	Shoulder ROM [deg]	31.211 \pm 4.338	30.116 \pm 4.593	0.911 [0.762, 0.965]

*Bold ICC values are interpreted as excellent (> 0.75).

Table 15: Test-retest reliability of reaching kinematics average *between* Day 1 and Day 2.

Parameters	Day 1	Day 2	ICC [95% CI]
Path Length [m]	0.368 \pm 0.052	0.357 \pm 0.059	0.835 [0.593, 0.934]
Movement Time [s]	0.573 \pm 0.209	0.523 \pm 0.244	0.781 [0.459, 0.912]
Movement Units [no units]	1.115 \pm 0.087	1.135 \pm 0.105	0.729 [0.325, 0.892]
Average Speed [m/s]	0.801 \pm 0.285	0.887 \pm 0.362	0.791 [0.375, 0.966]
Elbow ROM [deg]	4.013 \pm 1.397	4.744 \pm 2.708	0.692 [0.254, 0.876]
Shoulder ROM [deg]	30.743 \pm 3.834	30.792 \pm 4.270	0.820 [0.539, 0.929]

*Bold ICC values are interpreted as excellent (> 0.75).

0.979), except for the elbow ROM in Day 1, and movement units in both days (Table ??). Our ICC values were consistent with those reported in the literature using the Kinect [?]. Similarly, the test-retest reliability between Day 1 and Day 2 was excellent for path length, movement time, average speed, and shoulder ROM (from **0.729** to **0.835**), and it was fair to good for movement units and elbow ROM (from **0.692** to **0.729**) (Table ??). These results are consistent with the one study reporting test-retest reliability between days ($ICC > 0.66$ for all raters) [?]. For more details on the results relative to each arm independently refer to [?].

The study discusses several limitation in [?]. First, all participants belong to the same demographics (able-bodied adults), which might have created imperfect test-retest reliability between days. Future studies should increase the heterogeneity of

participants by including different age groups (e.g. young children) and/or clinical populations (e.g. children with cerebral palsy). Another limitation is that, when examining the within-day reliability (i.e. Table ??), there were more fair-to-good instances on Day 1 than on Day 2. Although participants had enough time to practice playing the game before the data collection process, it may be the case that there is a learning effect that occurred in Day 1 affecting our between-session reliability. As such, future studies should add more time to practice.

5.4 *Summary*

This chapter validated the efficiency of the *Super Pop VRTM* game by evaluating its accuracy and reliability relative to its sensing methods and measurements of reaching kinematics. The pilot study described in Section ?? evaluated the KinectTM camera's performance with respect to the OptiTrack, a current state-of-the-art motion capture system. Results show that 1) both motion capture systems yield similar outcome measures when evaluating users' reaching kinematics with the *Super Pop VRTM* game, and 2) the trajectories generated by both tracking systems are spatially similar to each other. These results support the validity of using the Kinect as the motion capture system for home-based rehabilitation purposes.

Section ?? discussed the testing sessions conducted to determine the *Super Pop VRTM* game's test-retest reliability within and between days. After analyzing the intraclass correlation coefficients (ICC) computed relative to a set of kinematic parameters, the study concludes that the game has good to excellent reliability between sessions and between days for most parameters. These results serve as an extension of the study described in Section ??, further confirming the potential for using the *Super Pop VRTM* game as a reliable evaluation tool to measure individuals' reaching kinematics.

CHAPTER VI

UPPER-BODY MOVEMENT CLASSIFICATION AND BASELINE SELECTION

6.1 Introduction

When providing targeted corrective feedback as part of physiotherapy intervention protocols, it's important to make sure that the complexity of the provided feedback cues is relative not only to the individual's needs, but also to the level of the individual's abilities. If the provided feedback cues require the individual to perform at a level that is significantly below his/her abilities, he/she might become bored due to the simplicity of the task. Similarly, if the provided feedback cues require the individual to perform at a level that is significantly above his/her abilities, he/she might become frustrated due to the task being too challenging. Neither case presents an efficient manner to achieve the desired effects.

To enable our system to provide corrective feedback in an effective and efficient manner, it first must have the capability to identify the level of the individual's abilities. We aim to achieve this by training a pattern recognition classification model with examples of typical and non-typical upper-body movement profiles relative to their reaching kinematics. Thus, at each iteration of completing a reaching task, our system can classify the user's performance and identify his/her **kinematic level**, which we define as the user's ability to complete a reaching task with respect to a set of kinematic parameters.

In the context of our system, providing feedback relative to the level of the user's abilities refers to selecting the baseline model (i.e. the model against which the user's kinematic performance is compared to) that is closest to the user's kinematic level.

By selecting an appropriate baseline model, we avoid providing feedback cues that would require the user to perform at a level that is significantly different than his/her abilities, thus preventing the user from becoming bored or frustrated.

This chapter describes the approaches taken to train and test different pattern recognition methodologies. The main purpose of the final classification model is to distinguish between upper-body movement profiles from different demographics as a function of the movement’s reaching kinematics. Section ?? describes the classes of interest and the descriptive features used to train the classification models. Section ?? discusses the approach we took to scale the data to reduce the potential bias introduced by the different value ranges of each feature. Given the nature of our collected data, we chose to experiment with the kMeans clustering algorithm, Gaussian Discriminant Analysis, and Support Vector Machines with three different kernel functions: linear, polynomial, and radial basis function. The general description of each methodology and the measures used to evaluate their performance are described in Section ?. Section ? describes the Majority Voting and Thresholding (MVTA) algorithm we developed to examine the ability of the trained classification models to distinguish between individuals as a whole instead of individual instances. Finally, Section ? discusses the results obtained from training and testing the trained classifiers on different data sets (instance-based classification), and those obtained from implementing the MVTA (individual-based classification).

6.2 Data Acquisition and Processing

6.2.1 Class Description

To construct our pattern recognition classification models, we recruited volunteers from different demographics (i.e. age, diagnosis, body dimensions, and kinematic abilities) to interact with our *Super Pop VRTM* system. They each performed the

90° reaching task shown in Figure ?? multiple times. Although the data were collected throughout various testing sessions, the game settings remained constant for all (Appendix ??). The reaching kinematics of each completed task were computed relative to the parameters defined in Section ??, and each is considered as an instance belonging to the group/class that corresponds to the participant who completed the task.

Although the ultimate goal is for the system to be able to identify the user’s kinematic level with as much precision as possible, we first make a general distinction between four main groups/classes relative to their age and abilities. The groups/classes, abbreviations, number of participants, and number of instances for each are organized in Table ?. The number of instances are shown with respect to the participants’ ‘affected’ arm. The ‘affected’ arm refers to the non-dominant arm for typically developing children and able-bodied adults, and to the arm that reflects the most symptoms for the children who have cerebral palsy. We focus on the most affected arm because it’s the one that yields the biggest difference in reaching kinematics between classes.

Table 16: Main classes for constructing pattern recognition classifiers.

Classes	Abbreviation	Age Range [years]	No. of Participants	No. of Instances
Able-bodied Adults	<i>A</i>	> 18	22	860
Typically Developing Teenagers	<i>T</i>	15-17	34	152
Typically Developing Children	<i>C_{TD}</i>	7-12	40	750
Children with Cerebral Palsy	<i>C_{CP}</i>	7-12	10	335

6.2.2 Descriptive Features

The seven kinematic parameters defined in Section ?? are the descriptive features for the pattern recognition classifiers: deviation from path (DfP), elbow and shoulder range of motion (EROM and SROM respectively), path length (PL), movement time (MT), movement smoothness (MUs), and average speed (AvgS). We performed two-tailed unpaired *t*-tests on the reaching kinematics between each and all classes to examine how they differ from each other with respect to the descriptive features. All

tests were performed at a 95% confidence level and are of the null-hypothesis that, for a given feature, the values between two given classes come from distributions with equal means (i.e. are not statistically different from each other). The parameters for which we can't reject the null-hypothesis are shown in Table ??, together with their corresponding p -values in parentheses.

Table 17: Kinematic parameters for which we can't reject the t -tests null-hypotheses between the corresponding two classes (with their p -values).

Classes	T	C_{TD}	C_{CP}
A	AvgS (0.49)	SROM (0.65)	None*
T	-	MT (0.62) SROM (0.25)	SROM (0.16) MUs (0.08)
C_{TD}	-	-	None*

* The corresponding two classes are statistically different from each other with respect to all parameters.

The A vs C_{CP} and C_{TD} vs C_{CP} comparisons yield p -values less than 0.05 for all parameters. As such, for these pairs of classes we conclude that they are statistically different from each other with respect to all kinematic parameters. This observation suggests that these pairs of classes are most likely separable and a classification model that differentiates between them could potentially yield relatively high classification accuracy rates. Taking these three classes into consideration, we focus on solving the binary classification problem of distinguishing between two classes: upper-body movement profiles with characteristics that pertain to children who have cerebral palsy, and movement profiles without said characteristics. Thus, we examine the ability of different pattern recognition classifiers to distinguish between the C_{CP} and C_{TD} classes, and the C_{CP} and A classes. We made sure that all parameters are linearly independent from each other relative to the C_{CP} vs C_{TD} and C_{CP} vs A pairs by confirming that their corresponding data sets are full rank.

6.2.3 Training and Test Sets

For any pattern recognition classification methodology, it's highly desirable for it to generalize to data not seen during the training process such that it can accurately classify unknown data [?]. As such, we avoid over-fitting the classification models by separating the data into training and test sets. We train the classification models and tune the corresponding model parameters with the training set, and test its overall performance with the test set. The classification accuracy rates obtained from the test set more precisely reflects the model's overall performance on classifying an unknown data set.

When comparing the different classification methodologies and the different approaches in manipulating the data, we make sure that the classification accuracy rates are comparable to each other by training and testing each with the same training and test sets. Considering the three groups of interest (C_{CP} , C_{TD} , and A), we extracted 25% of all 1,945 instances at random to construct the test sets. The remaining 75% were used to construct the training sets. More specifically, 69% of the C_{CP} vs C_{TD} data set belongs to the C_{TD} class and the remaining 31% belongs to the C_{CP} class. Similarly, 72% of the C_{CP} vs A data set belongs to the A class and the remaining 28% belongs to the C_{CP} class. These class distributions are maintained roughly the same for their corresponding training and test sets.

6.3 Feature Scaling: z-Score Standardization

The kinematic parameters defined in Section ?? describe different aspects of a given movement profile and they all have different units (i.e. millimeters, milliseconds, degrees, etc.), and thus are in different scales. Scaling the data before training any classification model offers the advantage of reducing the effect of features with greater numeric ranges dominating those with smaller numeric ranges [?]. As such, in addition to training and testing the pattern recognition methodologies of interest using the raw

data, we also examine the effects of scaling the data before training any classification method.

One way to standardize the values of the features into a specific range is by applying the **z-score standardization** method. This method has been shown to increase the performance of classifiers that aim to classify gait features that correspond to cerebral palsy movement profiles [?]. It has also been shown that the z-score standardization method is more effective and efficient than other methods, like the min-max and decimal scaling standardization method [?]. We apply this method to the data in the training set such that the values of each feature have zero-mean and unit-variance. For each descriptive feature, the standard scores of the corresponding values are computed with $\vec{x}' = (\vec{x} - \mu) / \sigma$; where \vec{x} is the vector containing the values of a given feature, and μ and σ are the mean and standard deviation of the training set's values for the corresponding feature. To ensure that the manipulation of the test set is the same as that of the training set, we scale the test set with the same μ and σ values.

We examine the effects of scaling the raw data by comparing the distribution of the feature values (Figure ??). The distribution of the feature values are represented with boxplots and are shown in Figure ?? and Figure ?? for the raw and scaled data respectively. We observe that the feature values of the raw data set are in widely different scales thus potentially adding a bias based on the corresponding range of values. For example, the movement time (MT) parameter has a range of 9,669 while the movement smoothness (MUs) parameter has a range of 14. This discrepancy in ranges may cause classifiers to give more weight to the feature with the higher range of values. By scaling the data we standardize the range of values for all features reducing the potential effect of their scales. This observation is supported by observing that all feature values of the scaled data have a range of values between 8-10 after applying the z-score standardization method.

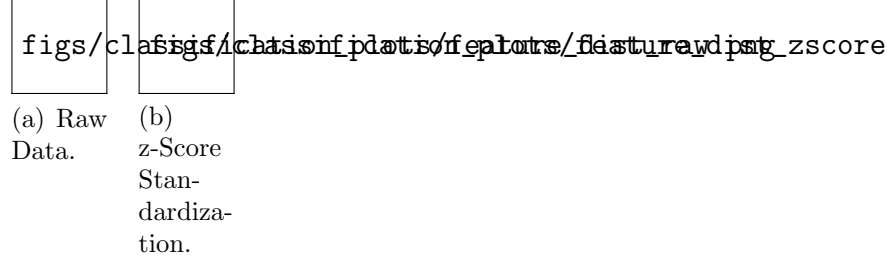


Figure 21: Effects of the different scaling approaches applied to the raw data.

In addition to reducing the potential effect of the different ranges in feature values, scaling the data also reduces the overlap between the classes of interest. Increasing the separation between the classes of interest allows for an increase in classification accuracy as it makes it easier for the classification models to distinguish between the classes. This observation is supported by examining the distance between class means. The normalized Euclidean distances between the means of the classes of interest (measured by scaling each feature by their maximum value), relative to the raw and scaled data are shown in Table ?? . As expected, the distances between the C_{CP} and C_{TD} classes, and between the C_{CP} and A classes, are larger when computed with the scaled data than with the raw data. To better examine the benefits of feature scaling, we examine the performance of the classification models by training and testing with raw and scaled data sets.

Table 18: Normalized Euclidean distances between class means relative to the raw and scaled data (no units).

	Raw Data	z-Score
C_{CP} vs C_{TD}	0.096	95.1
C_{CP} vs A	0.217	149.7

6.4 Classification Methodologies

Based on our literature review, we did not find any pattern recognition classification models trained for the purpose of classifying an individual’s upper-body movements as a function of their kinematic performance. We did find several research efforts

that focused on the detection and classification of movement profiles with respect to different temporal-spatial gait parameters. J. Kamruzzaman et al. [?] compares the performance of Support Vector Machines (SVMs), Multilayer Perceptron, and Linear Discriminant Analysis to detect and classify children who have cerebral palsy relative to temporal-spatial gait parameters. Results show that SVMs yield the largest overall classification accuracy of 96% (sensitivity = 94.32% and specificity = 100%). Similarly, R. Begg et al. [?] investigated the use of SVMs for the automatic recognition of gait changes due to aging. Results show that SVMs, using any one of six kernel functions, were able to achieve an overall accuracy of 91.7% in its capacity to distinguish between the two gait patterns.

In this chapter we aim to achieve similar accuracy rates in autonomously classifying upper-body movement profiles as belonging to children who have cerebral palsy versus typically developing children. The following subsections describe the different pattern recognition methodologies we experimented with. Section ?? describes the different measures used to evaluate the performance of the classification models, and Section ?? compares the performance of the different approaches.

6.4.1 kMeans

As a first pass towards determining how the data are spatially clustered with respect to the kinematic parameters of interest, we ran the training data through the kMeans clustering algorithm [?]. In general, the kMeans algorithm is an unsupervised learning method that, during the training process, assigns each instance in the data set to the cluster/class with the closest mean. After all instances are assigned, the algorithm updates the mean and covariance of each cluster relative to their new members. This process is repeated until there are no more updates in re-assigning the instances. The end result are the statistics (mean and covariance) of the resulting k clusters. A new data point is classified by assigning it to the nearest cluster as determined by

a distance metric (e.g. Euclidean distance). In this work, we initialize the clusters with the means of each corresponding class relative to the data in the training set. Given that the performance of the kMeans clustering algorithm is a function of the starting points, additional research will need to be conducted to find an approach that autonomously selects the optimal starting points (i.e. that yield the highest classification accuracy rate).

We use the data set's true labels to compute the classification accuracy of the kMeans algorithm using the test data, which was not seen by the algorithm during the training process. We classify each instance by assigning it to the cluster as a function of the cluster means obtained from the trained kMeans model.

6.4.2 Gaussian Discriminant Analysis

Gaussian discriminant analysis (GDA) is a member of the generative learning algorithm (GLA) family [?, ?]. GLAs model each class independently as opposed to discriminative learning algorithms which try to learn the conditional distribution of the classes given a set of data points. For the binary classification problem, let D be a training set defined as $D = \{(\mathbf{x}_i, y_i)\}_{i=1}^L$; where each instance vector $\mathbf{x}_i \in \mathbb{R}^n$ is associated with output $y_i \in \{0, 1\}$. Then $p(\mathbf{x}|\mathbf{y} = 0)$ models the distribution of the negative class's features, and $p(\mathbf{x}|\mathbf{y} = 1)$ models the distribution of the positive class's features. When using GDA, we assume that $p(x|y)$ is distributed according to a multivariate normal distribution which is parameterized by a mean vector $\mu \in \mathbb{R}^n$, and a covariance matrix $\Sigma \in \mathbb{R}^{n \times n}$, where $\Sigma \geq 0$ is symmetric and positive semi-definite. Thus, the probability density function of each class $c \in \{0, 1\}$ is given by (??),

$$p(x; \mu_c, \Sigma_c) = \mathcal{N}(\mu_c, \Sigma_c) = \frac{1}{(2\pi)^{n/2} |\Sigma_c|^{1/2}} \exp\left(-\frac{1}{2}(x - \mu_c)^T \Sigma_c^{-1} (x - \mu_c)\right) \quad (28)$$

where the maximum likelihood estimate of parameters (μ_0, Σ_0) and (μ_1, Σ_1) are obtained by maximizing the log-likelihood of the data with respect to the parameters

of the model (??).

$$\mu_0 = \frac{\sum_{i=1}^L \mathbb{1}\{y_i = 0\} \mathbf{x}_i}{\sum_{i=1}^L \mathbb{1}\{y_i = 0\}} \quad (29a)$$

$$\mu_1 = \frac{\sum_{i=1}^L \mathbb{1}\{y_i = 1\} \mathbf{x}_i}{\sum_{i=1}^L \mathbb{1}\{y_i = 1\}} \quad (29b)$$

$$\Sigma_0 = \Sigma_1 = \frac{1}{L} \sum_{i=1}^L (\mathbf{x}_i - \mu_{y_i})(\mathbf{x}_i - \mu_{y_i})^T \quad (29c)$$

and $\mathbb{1}_A(x)$ is the indicator function:

$$\mathbb{1}_A(x) = \begin{cases} 1, & \text{if } x \in A \\ 0, & \text{otherwise} \end{cases} \quad (30)$$

After modeling $p(\mathbf{x}|\mathbf{y})$ and assuming some class priors $p(\mathbf{y})$ (for this work we assume equal priors), any new data point $\hat{\mathbf{x}} \in \mathbb{R}^n$ is classified by selecting the class that yields the greater posterior probability (??).

$$y = \arg \max_y p(\mathbf{y}|\hat{\mathbf{x}}) = \arg \max_y p(\hat{\mathbf{x}}|\mathbf{y}) p(\mathbf{y}) \quad (31)$$

6.4.3 Support Vector Machines

Support Vector Machines (SVMs) is a supervised learning method that separates the instances from a given data set by finding an optimal separating hyperplane (OSH) [?], (i.e. the one that yields the maximum distance from the nearest training instances). SVMs have demonstrated good generalization performance in face recognition [?], text categorization [?], and optical character recognition [?]. They have also been shown to yield excellent classification accuracy rates in identifying gait patterns that belong to individuals with cerebral palsy [?, ?].

For binary classification problems, given that the original feature space can be too restrictive to search for the OSH, SVMs transform the training data into a higher

dimensional space for the construction of the OSH. Given a kernel function, SVMs perform the nonlinear mapping and construct a linear hyperplane between the two classes in the new feature space.

Let D be a training set defined as $D = \{(\mathbf{x}_i, y_i)\}_{i=1}^L$; where each instance vector $\mathbf{x}_i \in \mathbb{R}^n$ is associated with output $y_i \in \{-1, +1\}$. Let ϕ be the function that maps the training data into a higher dimensional space \mathcal{F} via nonlinear mapping $\phi : \mathbb{R}^n \rightarrow \mathcal{F}$. In the case where the data are linearly nonseparable in \mathcal{F} , there exists a vector $\mathbf{w} \in \mathcal{F}$ and a scalar b that define the separating hyperplane as $\mathbf{w}^T \phi(\mathbf{x}_i) + b = 0$. In general, determining the OSH that minimizes the training error and optimally separates the training data set in the higher dimensional space \mathcal{F} refers to solving the optimization problem (??).

$$\begin{aligned} \min_{\mathbf{w}, b, \xi} \quad & \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{i=1}^L \xi_i \\ \text{subject to} \quad & y_i(\mathbf{w}^T \phi(\mathbf{x}_i) + b) \geq 1 - \xi_i, \\ & \xi_i \geq 0 \quad \forall i. \end{aligned} \tag{32}$$

where ξ_i are the slack variables which yield nonzero values only for mis-classified data points, and $C > 0$ is the penalty parameter of the error term. This term is known as the **regularization parameter** as it regulates the generalization performance of the trained model. Solving the optimization problem yields weight vector \mathbf{w} and scalar b such that any new data point $\mathbf{x} \in \mathbb{R}^n$ is classified by (??).

$$y = f(\mathbf{x}) = \text{sign}(\mathbf{w}^T \phi(\mathbf{x}) + b) \tag{33}$$

Without having to define ϕ explicitly, we can define a Kernel function that describes an inner product in the space \mathcal{F} : $K(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_i) \cdot \phi(\mathbf{x}_j)$. For this work, we experimented with the following kernels and compared the performances of the

resulting SVM models with each other and with the rest of the classification methodologies:

- **Linear:** $K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i^T \mathbf{x}_j$
- **Polynomial:** $K(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i^T \mathbf{x}_j + 1)^p$
- **Radial Basis Function (RBF):** $K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(\frac{-\|\mathbf{x}_i - \mathbf{x}_j\|^2}{\sigma^2}\right)$

where p and σ are the degree of the polynomial and the width of the RBF function, respectively. These kernel parameters, together with the regularization parameter C , are tuned during the training process. The influence of these parameter values on the overall accuracy of the classification models is discussed in Section ??.

6.4.4 Classifier Performance Measures

Among the various measures for assessing the performance of a given classifier, for our binary classification problem we use: accuracy, sensitivity (true positive rate), and specificity (true negative rate) [?, ?]. In addition, we also evaluated the models' classification accuracy rates relative to their F_1 scores, which measures a model's accuracy with respect to the test's precision and recall [?]. The following are the general definitions and equations of the used measures:

- *Accuracy:* indicates the overall detection accuracy for both classes.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \times 100\% \quad (34)$$

- *Sensitivity (true positive rate):* ability of the classifier to accurately classify the positive class (i.e. the CP class).

$$Sensitivity = \frac{TP}{TP + FN} \times 100\% \quad (35)$$

- *Specificity (true negative rate)*: ability of the classifier not to generate a false detection.

$$Specificity = \frac{TN}{TN + FP} \times 100\% \quad (36)$$

- *F₁ Score*: measures the classifier's accuracy with respect to its precision and accuracy.

$$F_1 = 2 \times \frac{precision \times recall}{precision + recall} \quad (37)$$

where $precision = \frac{TP}{TP+FP}$, and $recall = \frac{TP}{TP+FN}$.

where TP and FP are the number of true and false positives respectively (i.e. when the classifier identifies the positive/CP class), and TN and FN are the number of true and false negatives respectively (i.e. when the classifier identifies the negative/Not-CP class).

6.5 Majority Voting and Thresholding Algorithm

As previously mentioned, the reaching kinematics of each completed reaching task are considered as an instance belonging to the group that corresponds to the participant who completed the task. In addition to this instance-based classification approach, we also developed an individual-based classification approach: the **Majority Voting and Thresholding Algorithm (MVTA)**. The MVTA uses instance-based trained models to classify participants as a whole instead of focusing on individual instances of their reaching kinematics. It classifies a given participant by taking a majority vote on the resulting predictions of his/her instances of reaching kinematics. Moreover, our MVTA maximizes the overall performance of the trained classifiers by determining the optimal balance between the true and false accuracy rates in classifying the positive class (i.e. participants that belong to the class of children with cerebral palsy). The complete process of the MVTA is described by Algorithm ???. The following are the definitions of the functions used in the algorithm:

- $\text{classify}(X^{(j)}, TH^{(i)})$:

Function of any given trained classification model that classifies the m instances in $X^{(j)}$ relative to the threshold value $TH^{(i)}$.

- $\text{vote}(x)$:

The majority voting function applied to $x_k \in \{0, 1\} \forall k = 1, \dots, m$:

$$\text{vote}(x) = \begin{cases} 1, & \sum_{k=1}^m \mathbb{1}(x_k = 1) > \sum_{k=1}^m \mathbb{1}(x_k = 0) \\ 0, & \text{otherwise} \end{cases} \quad (38)$$

where $\mathbb{1}_A(x)$ is the indicator function (??).

- $\text{getF}_1\text{Score}(v, w)$:

Computes the F_1 score between the predicted labels in v and the known labels in w using (??).

Algorithm 3 *Majority Voting and Thresholding Algorithm*: Determine optimal threshold value and classify an individual relative to a set of reaching kinematic instances.

Require: $TH^{(i)}$: i^{th} threshold value ($\forall i = [1, \dots, T]$), $\mathbf{X}^{(j)}$: matrix with m instances from the j^{th} participant ($\forall j = [1, \dots, P]$), and $\mathbf{y}^{(j)} \in \{0, 1\}$: class label of the j^{th} participant

```

1: for  $i = 1 : T$  do
2:   for  $j = 1 : P$  do
3:      $x \leftarrow \text{classify}(X^{(j)}, TH^{(i)})$   $\triangleright x_k \in \{0, 1\} \forall k = 1, \dots, m$ 
4:      $idx^{(j)} \leftarrow \text{vote}(x)$   $\triangleright idx^{(j)}$ : class prediction of the  $j^{th}$  participant
5:   end for
6:    $F^{(i)} \leftarrow \text{getF}_1\text{Score}(idx, y)$   $\triangleright F^{(i)}$ :  $F_1$  score of the predictions in  $idx^{(j)}$ 
7: end for
8:  $TH_{op} \leftarrow$  optimal threshold value that yields the maximum  $F_1$  score out of  $F$ 
9: for  $j = 1 : P$  do
10:   $x \leftarrow \text{classify}(X^{(j)}, TH_{op})$ 
11:   $idx_f^{(j)} \leftarrow \text{vote}(x)$ 
12: end for
13: return  $idx_f$ 

```

Considering the full data set described in Section ??, our training and testing sets are composed of instances from 10 children who have cerebral palsy, 40 typically developing children, and 22 able-bodied adults, each with varying amounts of reaching kinematic instances (Table ??). Using the already trained models, we classify the participants relative to the C_{CP} vs C_{TD} and C_{CP} vs A class pairs. (Refer to Section ?? for the results and analysis of implementing the MVTA).

6.6 *Classification Performance Results and Discussion*

In this section we mainly investigate: 1) the ability of the methodologies described in Section ?? to distinguish between movement profiles belonging to children who have cerebral palsy (C_{CP}) versus typically developing children (C_{TD}), as well as between profiles belonging to children who have cerebral palsy versus adults (A), and 2) the effectiveness of our Majority Voting and Thresholding Algorithm (Section ??), to classify an individual as a whole instead of individual instances of reaching kinematics. In addition to the overall discussions on these instance- and individual-based classification approaches, this section also discusses the effectiveness of the z-score standardization method towards reducing the effect of different ranges in feature values when implementing trained classifiers (Section ??), and the influence of the regularization and kernel parameters on the overall accuracy of the Support Vector Machines (Section ??).

6.6.1 Quantitative Results: Instance-based Classification

The accuracy (??), sensitivity (??), and specificity rates (??), and the F_1 scores obtained from testing the models on seen and unseen raw and scaled data are shown in Table ?. The overall performance of a given classification model is determined based on the F_1 scores (as computed by (??)). All F_1 scores obtained from distinguishing the classes of interest with raw and scaled test/unseen data are shown in Figure ?.

Table 19: Accuracy, sensitivity, and specificity rates, and F_1 scores of different classifiers with raw and scaled reaching kinematics of the corresponding training and test sets. (All values are percentages).

Classifiers (kernel)	Performance Measures	C_{CP} vs C_{TD}				C_{CP} vs A			
		Raw Data		z-Score		Raw Data		z-Score	
		Train	Test	Train	Test	Train	Test	Train	Test
kMeans	Accuracy	68.5	69.9	66.3	66.2	76.5	78.3	77.8	79.6
	Sensitivity	20.4	22.4	24.8	30.6	21.2	23.5	24.0	30.6
	Specificity	89.9	91.4	84.7	82.4	97.8	100.0	98.6	99.1
	F_1 Score	28.5	31.7	31.2	36.1	33.4	38.1	37.6	46.0
GDA	Accuracy	68.0	66.9	68.0	66.9	84.8	89.0	84.8	89.0
	Sensitivity	52.5	68.2	52.8	68.2	71.6	81.2	71.6	81.2
	Specificity	74.8	66.3	74.8	66.3	89.9	92.1	89.9	92.1
	F_1 Score	50.4	56.3	50.4	56.3	72.5	80.7	72.5	80.7
SVM (linear)	Accuracy	69.7	66.5	70.5	66.9	83.6	89.0	82.9	88.0
	Sensitivity	55.6	63.5	53.6	62.4	77.2	85.9	77.6	85.9
	Specificity	76.0	67.9	78.0	69.0	86.1	90.2	85.0	88.8
	F_1 Score	53.1	54.3	52.8	54.1	72.4	81.6	71.7	80.2
SVM (polynomial)	Accuracy	69.7	66.2	73.7	69.9	83.3	88.6	96.2	90.6
	Sensitivity	53.2	61.2	73.6	78.8	77.6	85.9	94.8	78.8
	Specificity	77.1	68.4	73.7	65.8	85.4	89.7	96.7	95.3
	F_1 Score	52.0	53.1	63.2	62.0	72.1	81.1	93.3	82.7
SVM (RBF)	Accuracy	-	-	77.6	68.0	-	-	98.3	93.3
	Sensitivity	-	-	74.4	67.1	-	-	97.6	89.4
	Specificity	-	-	79.0	68.4	-	-	98.6	94.9
	F_1 Score	-	-	67.1	56.7	-	-	97.0	88.4

* Bold values refer to the model that yields the best performance for the corresponding class pair overall.

** Missing values reflect the poor ability of the corresponding model to distinguish between the classes of interest.

figs/classification_plots/all_F1_scores.png

Figure 22: Bar plots showing the F_1 scores obtained from classifying raw and scaled test data.

Our first observation from Table ?? and Figure ?? is that most classifiers yield greater or equal F_1 scores when trained and tested with scaled data than with raw data. Among all pattern recognition methodologies, for the C_{CP} vs C_{TD} comparison, the maximum F_1 score obtained with raw data is 56.3% versus 62.0% obtained with scaled data. Similarly for the C_{CP} vs A comparison, the maximum F_1 score obtained with raw data is 81.1% versus 88.4% obtained with scaled data. These results are supported by our analysis of the Euclidean distances between the classes of interest discussed in Section ?. For both class pairs, the normalized Euclidean distances between class means are larger when applying the z-score standardization method (Table ?). The farther away two classes are from each other, the less the overlap between them. As such, it stands to reason that it is easier for the trained classifiers to distinguish between movement profiles when the data is scaled.

Our second observation from Table ?? and Figure ?? is that all classifiers yield larger F_1 scores when distinguishing between the C_{CP} and A classes than between the C_{CP} and C_{TD} classes with both raw and scaled data. Among all pattern recognition methodologies, when using raw data, the maximum F_1 score obtained when comparing the C_{CP} and C_{TD} classes is 53.1% versus 72.5% obtained when comparing the C_{CP} and A classes. Similarly when using scaled data, the maximum F_1 score obtained when comparing the C_{CP} and C_{TD} classes is 62.0% versus 88.4% obtained when comparing the C_{CP} and A classes. This observation is supported by analyzing the distances between the classes of interest (Table ?). The normalized Euclidean distance between the means of the C_{CP} and A classes is greater than between the means of the C_{CP} and C_{TD} classes. As expected, these results suggest that the difference in movement profile characteristics, relative to the selected descriptive features, is more noticeable between children who have cerebral palsy and adults than it is between children who have cerebral palsy and typically developing children. Similar to the first observation, it stands to reason that it is easier for trained classifiers to

distinguish between classes that have less overlap between them. We suspect that this is the case because, in general, adults do a better job at following the protocol when it comes to completing the reaching task and thus make less mistakes that could be identified as characteristics belonging to movement profiles of children who have cerebral palsy.

In general, all methodologies do a fairly good job at generalizing the performance of the trained classifiers on unseen data. This is to say, the F_1 scores obtained from classifying unseen data are greater or similar to those obtained from classifying seen data. As expected, the kMeans clustering algorithm does a poor job at distinguishing the classes of interest (sensitivity = 30.6% for both class comparisons). Given that the classes of interest are too close to each other, classifying instances based on how close they are to the class means is prone to yielding large classification error rates. Estimating the conditional distribution of the classes and classifying instances to the class that yields the greater posterior probability is a better approach towards distinguishing between the classes of interest, but the large amount of overlap between them is still a concern. When distinguishing between the C_{CP} and C_{TD} classes, Gaussian Discriminant Analysis (GDA) yields sensitivity and specificity rates of 68.2% and 66.3% respectively. When distinguishing between the C_{CP} and A classes, GDA yields sensitivity and specificity rates of 81.2% and 92.1% respectively. These results are expected given that the normalized Euclidean distance between the means of the C_{CP} and A classes is greater than that between the means of the C_{CP} and C_{TD} classes (Table ??).

Finally, for the C_{CP} vs C_{TD} comparison, the classifier that yields the best overall performance is obtained with the SVM methodology using a polynomial kernel function. When tested with unseen data, the trained model yields a classification accuracy rate of 69.9% and is capable of identifying movement profiles with characteristics of children who have cerebral palsy with a fairly high accuracy rate of 78.8%. Similarly,

for the C_{CP} vs A comparison, the classifier that yields the best overall performance is obtained with the SVM methodology using a radial basis kernel function. When tested with unseen data, the trained model yield a classification accuracy rate of 93.3% and is capable of identifying movement profiles with characteristics of children who have cerebral palsy with a significantly high accuracy rate of 89.4%.

6.6.1.1 Kernel Parameter Selection

The results shown in Table ?? for the SVMs were obtained by training models with the optimal values for the parameters of the corresponding kernel functions (i.e. the values that yield the best overall performance). The parameters that were tuned during the training process were the regularization parameter for all kernel functions, the degree of the polynomial kernel function, and the width of the radial basis function (Section ??). For each kernel function, we used 10-fold cross-validation on the corresponding training set to test different values for all kernel parameters and select the combination that can best classify unknown data. We adhere to the suggestions from [?]: the polynomial degree (p) and the width of the radial basis function (σ) were varied from 1 to 20 and from 0.01 to 2.5, respectively. For each p and σ , the value of the regularization parameter (C) was varied from 10^{-5} to 10^5 . For each classifier, we select the combination parameter values that yield the greater F_1 scores.

We take into consideration the classifiers that yield the best overall performance on scaled unseen data for both class comparisons. The influence of the kernel parameters on the overall performance of the classifiers is shown in Figure ?? for the SVM using a polynomial kernel function (C_{CP} vs C_{TD} comparison), and in Figure ?? for the SVM using a radial basis kernel function (C_{CP} vs A comparison). The F_1 scores shown in Figure ?? and Figure ?? are relative to the polynomial degree and the value of the regularization parameter, respectively, that yields the greater F_1 scores. The same applies to the case of using the radial basis kernel function in Figure ??.

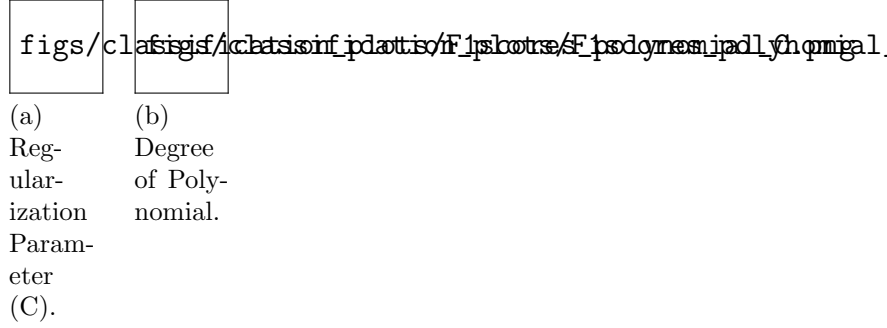


Figure 23: F_1 scores as a function of the regularization parameter and the degree of the polynomial kernel function for the C_{CP} vs C_{TD} comparison.

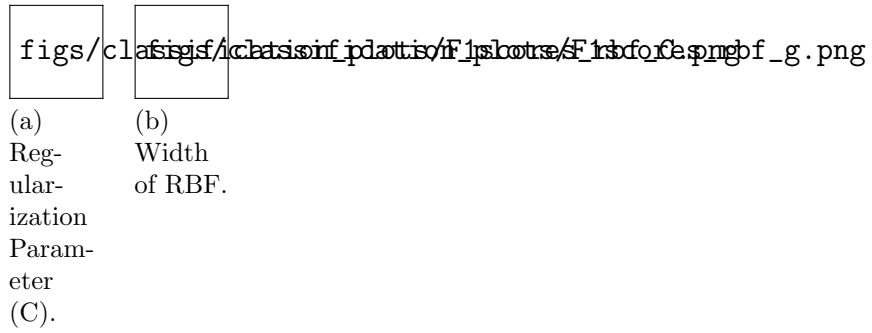


Figure 24: F_1 scores as a function of the regularization parameter and the width of the radial basis function for the C_{CP} vs A comparison.

6.6.2 Quantitative Results: Individual-based Classification (MVTA)

This section describes the results obtained from testing our Majority Voting and Thresholding Algorithm (MVTA), described in Section ??, as an individual-based classification approach. We test the MVTA using the instance-based trained classifiers that yield the best overall performance for the two class pairs. For SVMs, we first compute the scores of a set of instances. The SVM score for classifying an instance x_i is the signed distance from x_i to the trained decision boundary. If the score of a given instance is greater than or equal to the threshold value, the instance is classified as a member of the positive class (i.e. C_{CP}). If the score is less than the threshold value, the instance is classified as a member of the negative class (i.e. C_{TD} or A).

The accuracy, sensitivity, and specificity rates, and the F_1 scores resulting from testing the MVTA on both class pairs are shown in Table ??. The F_1 scores as a function of the tested threshold values for both class pairs are shown in Figure ??, as well as the maximum possible F_1 score for both cases. Overall, the MVTA improved the classifiers' performance on distinguishing between the classes of interest. There was a 21.2% and 4.2% increase in sensitivity and specificity rates, respectively, for the C_{CP} vs C_{TD} comparison (sensitivity = 100%; specificity = 70%); and a 10.6% and 5.1% increase in sensitivity and specificity rates, respectively, for the C_{CP} vs A comparison (sensitivity = 100%; specificity = 100%). As previously mentioned, we suspect that this is the case because adults do a better job at following the protocol when completing the required reaching tasks. This is to say that, although typically developing children do not have any symptoms that may affect their kinematic performance, they still make some mistakes when completing the required reaching task.

Table 20: Accuracy, sensitivity, and specificity rates, and F_1 scores from the Majority Voting and Thresholding approach. (All values are percentages).

Performance Measures	C_{CP} vs C_{TD}	C_{CP} vs A
	SVM (polynomial)	SVM (RBF)
Accuracy	76.0	100.0
Sensitivity	100.0	100.0
Specificity	70.0	100.0
F_1 Score	62.5	100.0

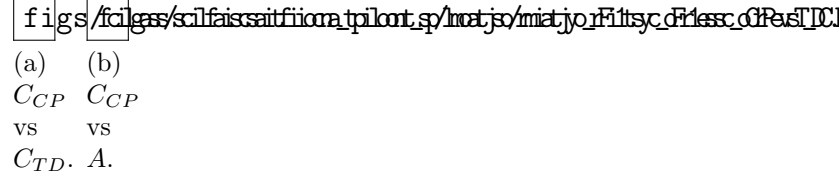


Figure 25: F_1 scores as a function of the different threshold values used for classifying participants as instances.

6.7 Summary

In the context of providing targeted corrective feedback as part of physiotherapy intervention protocols, we discuss the importance of providing the feedback cues relative to the level of the individual's abilities to minimize potential boredom or frustration. For our *Super Pop VRTM* system to provide corrective feedback in an efficient and effective manner, it first must have the capability to identify the level of the individual's abilities. We aim to achieve this by training a pattern recognition classification model that can classify users' upper-body movement profiles as a function of their reaching kinematics. In this manner, at each iteration of completing a reaching task, our system can identify the user's kinematic level (i.e. the ability to complete a reaching task with respect to a set of kinematic parameters) and select the baseline model that is closest to said level. By selecting an appropriate baseline model we avoid providing feedback cues that would require the user to perform at a level that is significantly different than his/her abilities, thus preventing the user from becoming bored or frustrated.

This chapter describes the testing conducted to train and test different pattern

recognition methodologies, and the different approaches taken to increase the models' overall performances. Section ?? describes the classes of interest and the descriptive features used to train the classification models. Section ?? discusses the scaling approaches we experimented with to reduce the potential bias introduced by different value ranges of the descriptive features. The effects of scaling the raw data is shown in Figure ?. Section ?? describes the three pattern recognition classification methodologies we experimented with: the kMeans clustering algorithm, Gaussian Discriminant Analysis, and Support Vector Machines (SVMs) with three different kernel functions: linear, polynomial, and radial basis function. In addition to testing the classifiers with an instance-based classification approach, we also discuss our Majority Voting and Threshold Algorithm as an individual-based classification approach in Section ?. Given that participants without cerebral palsy can make mistakes when completing the required reaching tasks, a few of their upper-body movement profile instances may contain characteristics that belong to the group of children who have cerebral palsy. Thus, the goal of the individual-based classification approach is to improve the classifiers' overall performance by classifying participants as a whole such that typically developing children and able-bodied adults will be correctly classified even if they do make a mistake when completing the required reaching task. Finally, Section ?? compares the performances of the trained models for each classification methodology relative to the performance measures described in Section ?. Taking into consideration the classifiers that yield the best performance for the two class pairs, the section also discusses the results of our individual-based classification approach.

Overall, the results of this study suggest that Support Vector Machines, using the polynomial and radial basis kernel functions and using the z-score standardization method for scaling the data, are capable of distinguishing between movement profiles that belong to children who have cerebral palsy versus typically developing children and between children who have cerebral palsy versus adults, respectively, as

a function of their upper-body reaching kinematics. Moreover, results confirm that it is easier for classifiers to distinguish between adults and children who have cerebral palsy, than between typically developing children and children who have cerebral palsy. This observation is supported by the results obtained from the individual-based classification approach. Although the classifiers' overall performance improved by implementing our MVTA, the overall performance on distinguishing between the C_{CP} and C_{TD} classes (F_1 score = 62.5%), is still lower than that on distinguishing between the C_{CP} and A classes (F_1 score = 100%).

The resulting classification models are the first step towards identifying a user's kinematic level such that our system can provide the necessary appropriate/optimal corrective feedback cues to effectively induce improvement in the user's kinematic performance. Moving forward, more data would need to be collected such that we can train classification models with the capability of better distinguishing between more specific levels of kinematic performance.

CHAPTER VII

KINEMATIC BEHAVIOR ADAPTATION VIA ROBOTIC INTERACTIONS

7.1 *Introduction*

We define an individual's **kinematic behavior** as their behavior/performance relative to a set of kinematic parameters. Previous studies have shown that external corrective feedback provided to individuals undergoing physical therapy sessions increases the efficacy of their intervention protocols by prompting them to modify their kinematic behavior, thus allowing for individuals to facilitate sustained or complex play [?]. (Refer to Section ?? for more details on the benefits of feedback). In traditional physical therapy sessions, an expert therapist is able to achieve desired results from their patient through 1) guided instruction, 2) thorough observation, 3) real-time assessment, and 4) corrective feedback. The therapist first provides the patient with guided instruction, which is the initial description of the task/movement that is required of the patient to perform. The therapist observes the patient perform the task, assesses the movement, and provides the patient with corrective feedback for improvement. The cycle is repeated until the desired results are reached. However, direct feedback is typically provided by an expert therapist during weekly or monthly visits, which limits improvement on a daily basis. As such, we promote in-home rehabilitation protocols via robot interactions while using our *Super Pop VRTM* system such that users can receive the necessary corrective feedback that prompts them to adapt their kinematic behavior to, ultimately, increase their rate of improvement.

This chapter focuses on the studies we conducted to support our claim that individuals can modify and adapt their kinematic behavior while interacting with our

complete *Super Pop VRTM* system. Our approach follows a procedure similar to the one shown in Figure ???. Instead of classifying the user’s reaching kinematics to autonomously select the most appropriate baseline model, the studies described in this chapter assume a constant one (Figure ??). Using this approach, we are able to constrain the testing protocol such that we attribute any kinematic behavior adaptation solely to the corrective feedback provided by the system.

The first study is described in Section ??. Results show that both, able-bodied adults and typically developing children, can reach specific performance reference values for a given kinematic parameter via receiving appropriate corrective feedback from a robotic playmate. The second study is described in Section ??. Results show that both typically developing children and children who have cerebral palsy can decrease their movement time after a training session with a robotic playmate. Both studies measure the participants’ outcome measures relative to the *Super Pop VRTM* environment.

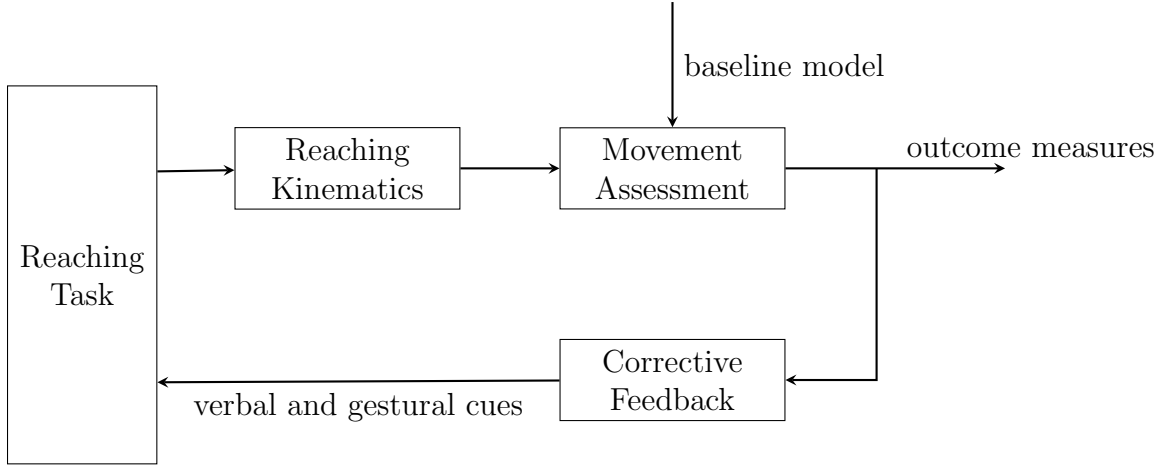


Figure 26: Diagram of our complete system with a constant baseline model.

7.1.1 DARwIn-OP: Humanoid Robot

The two studies described in this chapter make use of an embodied physical robotic playmate to provide feedback cues to assist users in reaching their kinematic goals.

Both studies use the humanoid robot, **DARwIn-OP** (Darwin) (Figure ??). It has 20 actuators resulting in 6 DOF (degrees of freedom) for each leg, 3 DOF for each arm, and 2 DOF for the neck [?]. Darwin was pre-programmed with a library of verbal and nonverbal behaviors to enable interaction with the Super Pop environment and provide feedback to the user [?]. The different combinations of these behaviors are summarized in the studies’ corresponding sections (?? and ??). In both studies, Darwin introduces himself with the script below to provide some low-level instructions.

“Hello. My name is Darwin, and I will be playing Super Pop with you today. I will ask you to complete a series of tasks, and I would love it if you would follow my instructions. When you’re ready, please raise both of your hands as high as you can.”

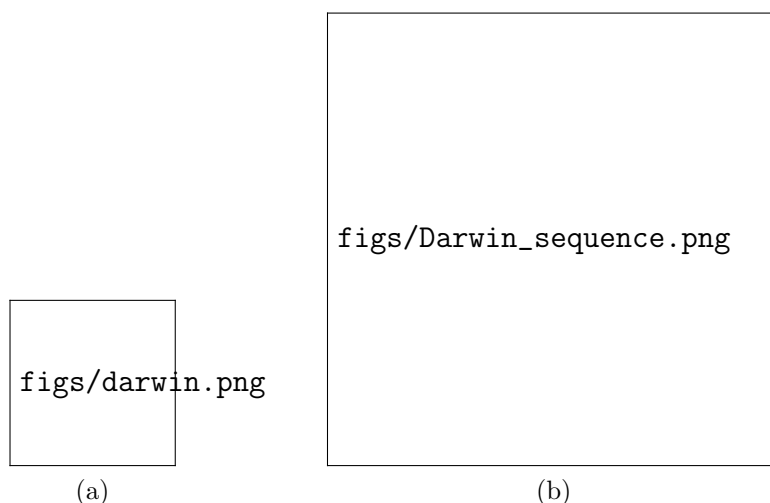


Figure 27: (a) Humanoid robot, DARwIn-OP (Darwin). (Image adapted from [?].) (b) Darwin performing a 90° reaching task.

7.1.2 Effect of Robot Feedback on Motor Skill Performance

We first conducted a pilot study to evaluate how various cues affect an individual’s kinematic performance [?]. More specifically, we employed a between-groups experimental design to compare the effect of verbal versus combined verbal and nonverbal

instructional feedback provided by the robotic playmate. The most effective combination of feedback cues (i.e. the combination that induced a greater rate of change in the participants' kinematic behavior) was selected to be used in the studies described in this chapter. For this pilot study, the participants' kinematic behavior was analyzed with respect to the movement time parameter. We tested the system using the procedure as shown Figure ?? with 20 able-bodied adults (5 females and 15 males ranging in age between 18 and 45 years old, mean = 28.4 years and standard deviation = 5.7 years). To assess the upper-body movements of adult participants with respect to the movement time parameter, the constant baseline model used for this study was the Fitts model described in Section ??.

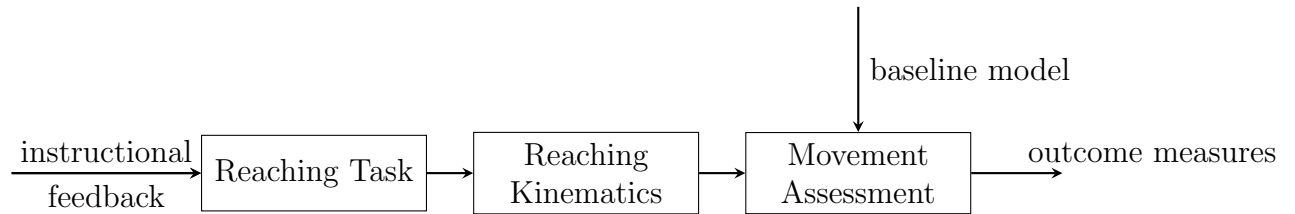


Figure 28: Diagram of our system with a constant baseline model and instructional feedback is received as input instead of corrective feedback.

Each participant was randomly assigned to one of two groups, and asked to perform the 90° exercise shown in Figure ?? as the reaching task, three times. Participants in Group 1 received verbal instructional feedback only, while participants in Group 2 received a combination of verbal and nonverbal instructional feedback. Participants in both groups received the following verbal instructions before each corresponding task: 1) move “*at a speed that feels normal*”, 2) move “*as slow as possible*”, and 3) move “*at a speed that is a little slower than normal*”. Only participants in Group 2 received, in addition, the nonverbal instructional cues from Darwin (Figure ??).

For each group and for Darwin, the normalized average movement times with respect to each task for Group 1, Group 2, and Darwin are shown in Figure ??.

Participants from both groups were able to follow Darwin’s instructions accordingly (i.e. both groups began at a certain baseline, were able to slow down for Task 2, and then speed up for Task 3). However, when examining Task 2 (Darwin instructs participants to ‘move as slow as possible’), the normalized average movement times were 0.38 ± 0.26 and 0.67 ± 0.30 for Group 1 and Group 2 respectively. Seeing as Group 2 performed closer to Darwin’s instructions (i.e. they took longer to perform the task), these results suggest that individuals receiving a combination of verbal and nonverbal feedback can perform closer to the robot’s guided instructions when compared to only receiving verbal instructional feedback. As such, we adhere to this study’s results and provide a **combination of verbal and nonverbal feedback cues** in the studies described in this chapter. For more detailed results, refer to [?].

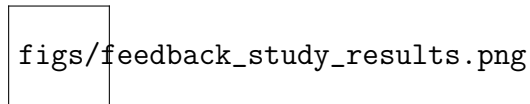


Figure 29: The normalized average movement times with respect to each task for Group 1, Group 2, and Darwin. (Image adapted from [?].)

7.2 *Pilot Study I: Low-resolution Feedback from Robotic Playmate*

An embodied robotic agent providing guided instruction to individuals during physical therapy sessions has been shown to improve individual kinematic performance [?] and increase overall engagement levels [?]. Given that the main objective of physical therapy with a human therapist is to reach a performance goal over an extended period of time, we aim to reach these long-term performance goals through implementation of a robotic playmate providing continuous feedback (Figure ??). As such, we conducted a study with the main purpose to determine if individuals can adapt their kinematic behavior via robot interaction while using our *Super Pop VRTM* system such that they improve their performance at each iteration of feedback and, ultimately, reach

an individualized performance goal [?].

7.2.1 Hypotheses

To better evaluate the efficacy of our system, we compare the effects of receiving feedback from the embodied robotic playmate versus the baseline effects of receiving feedback from a virtual agent. As such, the hypotheses of this study were:

H_1 : By interacting with our overall *Super Pop VRTM* system, users will improve their kinematic performance at each instance of completing a reaching task and, ultimately, reach a targeted performance goal.

H_2 : Users will reach their individualized performance goals faster, on average, when receiving feedback from the embodied robotic playmate than when receiving feedback from a virtual agent.

7.2.2 Experimental Design

Fifteen able-bodied adults and fourteen typically developing children were recruited to interact with our *Super Pop VRTM* system (Figure ??). There were six females and nine males in the adult group, ranging in age between 19 and 33 (mean age = 26.9 years, standard deviation = 3.4 years), and five females and nine males in the children group, ranging in age between 15 and 16 years (mean age = 15.5 years, standard deviation = 0.5 years). Adult participants and the parents of the child participants signed the IRB (Institutional Review Board) approved consent form allowing them to engage in the testing sessions.

This study focused on computing and correcting participants' movement time (MT) via feedback. As such, the targeted performance goal that participants were prompted to reach was the corresponding MT prediction computed by our Fitts model (refer to Section ??). For the first round of experiments, the adult participants were randomly assigned to one of two groups: **Group A** received feedback from a virtual

agent while **Group B** received feedback from the robotic playmate (refer to Section ?? for more details on Darwin). For the virtual agent, we played Darwin’s voice over external speakers. Both Darwin and the speakers were positions between the screen and the participant, and they both provided corrective feedback as described in Table ?. The purpose of the virtual agent is to establish a baseline comparison to evaluate whether an embodied agent can yield better results than a virtual agent (i.e. a virtual voice).

Table 21: Feedback provided during interaction with the *Super Pop VRTM* game.

Movement Time, MT	Verbal	Nonverbal
$MT > target$	“Great job. Move a little faster like this...”	Darwin performs the gesture at the correct movement time. It extends his left arm above his head. It then moves his shoulder joint 90° until his arm is down and parallel to his torso (Figure ??).
$MT < target$	“Great job. Move a little slower like this...”	
$MT = target$	“Fantastic.”	

**target* is defined as the MT reference window (i.e. the MT target value $\pm \epsilon$ to allow for small deviation from the actual MT target value).

All participants followed a variation of the protocol described in Appendix ?. Each participant interacted with the system for one round. Before the game starts, the system computes the user’s MT reference and a ± 150 ms margin of error is added to the final value. Each round consists of three main steps: 1) user performs the reaching task (i.e. the 90° exercise shown in Figure ??), 2) the system compares the user’s MT to the reference, and 3) either Darwin or the virtual agent provides the corresponding corrective feedback until the user reaches the goal or until the time runs out (6 minutes). The flowchart describing the architecture of the testing sessions is shown in Figure ?.

For all participants, at the beginning of each round we make sure that the participants’ starting MT is much greater than the reference MT by having the virtual voice or Darwin speak the following:

“Please pop bubbles one through three and move as slow as you can. Like

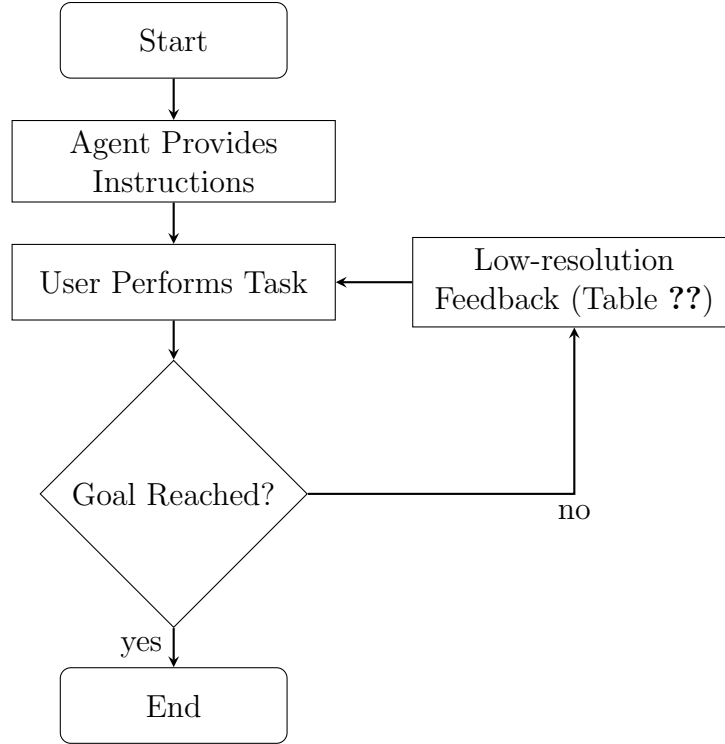


Figure 30: Flowchart describing the interaction between the user and the system.

this...”

For all sessions in Group B, Darwin also performs the nonverbal gesture that is 10 times slower than the one described in Table ?? to show the user how the reaching task should be completed. These instructions prompt the users to move as slow as they can thus defining a common starting point between all participants.

For the second round of experiments, child participants interacted with the version of the system that provides feedback via the robotic playmate to see if a younger demographic could also reach their performance goals by interacting with the system.

7.2.3 Quantitative Results (Adults)

All adult participants from both groups reached their corresponding MT references. For each participant, the number of trials children needed to reach their corresponding MT references, and their averages and standard deviations, are shown in Table ??

organized by groups. Moreover, boxplots showing how the number of trials needed to reach the MT references are distributed per group are shown in Figure ??.

Table 22: Trials to MT references for all participants from each group.

Participant	Adults (Virtual Agent)	Adults (Darwin)	Children (Darwin)
1	5	1	5
2	7	4	3
3	4	1	4
4	4	2	3
5	5	4	3
6	2	2	4
7	2	2	7
8	3	-	3
9	-	-	5
10	-	-	N/A*
11	-	-	4
12	-	-	2
13	-	-	1
14	-	-	8
AVG	4.0	2.3	4.0
STD	1.7	1.3	1.9

*Participant did not reach his/her MT reference.

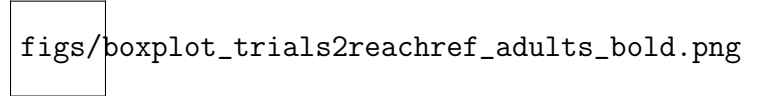


Figure 31: Boxplots showing how the number of trials needed to reach the MT references are distributed with respect to each group of adult participants.

7.2.4 Quantitative Results (Children)

Thirteen out of the 14 child participants reached their corresponding MT references. As an example, the response of Participant 4, who reached the MT reference in 3 trials, is shown in Figure ?. On the other hand, the response of Participant 10, who did not reach the MT reference, is shown in Figure ?. We discuss some of the potential reasons for this participant not reaching his MT reference in the *Discussion and Conclusions* section. The number of trials children needed to reach their corresponding MT references, and their averages and standard deviations are shown in Table ??, and a boxplot showing how the number of trials needed to reach the MT references are distributed is shown in Figure ?.

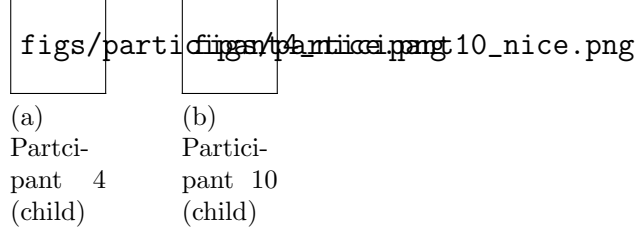


Figure 32: MT response curves of a) Participant 4 (child) who reached the MT reference at trial 3, and b) Participant 10 (child) who did not reach the MT reference.

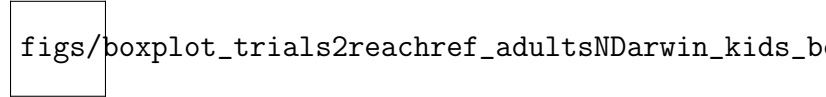


Figure 33: Boxplots comparing the number of trials needed to reach the MT references for child and adult participants.

7.2.5 Discussion and Conclusions

The first round of experiments consisted of comparing the responses between the two adult groups that received feedback from the virtual agent (Group A) and from the robotic playmate, Darwin (Group B). Group A and Group B needed an average of **4.0 ± 1.7 trials** and **2.3 ± 1.3 trials**, respectively, to reach their respective MT references (Table ??). These results, together with the boxplots in Figure ??, validate our second hypothesis (H_2): individuals that receive corrective feedback from a robotic agent (Group B) need, on average, less trials to reach their respective MT references than individuals that receive feedback from a virtual agent (Group A). We performed a two-tailed unpaired t -test at a 95% confidence level on the adults' results to determine if the difference between the groups' amount of trials needed to reach the MT references is statistically significant. We define our null-hypothesis as: the two independent samples come from distributions with equal means (i.e. the participants in the two groups needed, on average, the same amount of trials to reach their respective MT references). Our t -test analysis results in a **p-value = 0.046**. As such, we reject the null-hypothesis and confirm the statistical significance of the

difference between the number of trials needed to reach the respective MT references between the two adult groups.

Because the results support the robotic playmate over the virtual agent, the second round of experiments involved children interacting with the version of the system that provides feedback via the robotic playmate. Results show that 13 out of the 14 children participants reached their corresponding MT references (Table ??). Thus, we performed a two-tailed paired t -test analysis at a 99% confidence level on the child participants' results to determine if the amount of participants that reached their references is statistically significant. Let \vec{v} and \vec{w} be two vectors containing the absolute difference between the child participants' MTs and their corresponding MT reference for their first and last trials respectively, where each element belongs to a different participant. Let \vec{d} be the vector containing the difference between \vec{v} and \vec{w} (??).

$$\vec{d} = \begin{bmatrix} d_1 \\ d_2 \\ \dots \\ d_n \end{bmatrix} = \begin{bmatrix} |Ft_1 - Rt_1| \\ |Ft_2 - Rt_2| \\ \dots \\ |Ft_n - Rt_n| \end{bmatrix} - \begin{bmatrix} |Lt_1 - Rt_1| \\ |Lt_2 - Rt_2| \\ \dots \\ |Lt_n - Rt_n| \end{bmatrix} \quad (39)$$

where n is the number of the participants being analyzed, Ft_i and Lt_i are the i^{th} participant's MT in the first and last trials respectively, and Rt_i is the i^{th} participant's MT reference. Thus, we define our null-hypothesis as: the sample mean of \vec{d} is equal to zero (i.e. there is no statistical difference between \vec{v} and \vec{w}). Our t -test analysis on \vec{d} results in a **p-value** $\ll 0.01$. As such, we reject the null-hypothesis and conclude that there is a statistical difference between the participants' performance in their first and last trials. This suggests that the participant that did not reach their MT reference does not affect our claim that children will reach their performance goals by interacting with our system. This, together with the fact that all adult participants

reached their corresponding performance goals, validates our first hypothesis (H_1). After observing the session and response curve of the participant that did not reach their performance goal (Figure ??), we hypothesize that one of the reasons might be because the participant did not understand Darwin’s feedback cues.

Moreover, even though the child participants received corrective feedback via the robotic playmate, they performed at a lower level (i.e. greater MT average) than the adults that interacted with the same version of the system. We attribute this observation to the fact that, in general, children have slower movements than adults [?].

One of the limitations of this work is related to the demographic of the volunteers recruited to participate in this study. Namely, even though the target population for our system are children with motor limitations who are enrolled in some physical therapy protocol, we recruited able-bodied adults and typically developing children. On a similar note, the performance of the children participants was evaluated against the same Fitts model as the adults participants were (Section ??). As such, future studies will include sessions with participants of the target population, and participants will be compared against their corresponding baselines. Another limitation of this work is the number of kinematic parameters used to evaluate the participants’ performances. Further studies will be conducted with additional kinematic parameters as well as combinations of parameters.

7.3 Pilot Study II: Decreasing Users’ Movement Times

The purpose of this pilot study was to examine whether instructional and corrective feedback cues provided by a robotic playmate could improve the kinematic performance of typically developing children and children who have cerebral palsy. In general, one of the main objectives of a physical therapy session with a human therapist is to improve patient performance and maintain the performance after the session has

ended [?, ?]. As such, we designed our *Super Pop VRTM* system such that it can allow for such interactions in the home environment. In this manner, users can continuously interact with the system towards maintaining their improvement in performance.

7.3.1 Hypotheses

For this study, participants' kinematic performance is evaluated with respect to the movement time (MT) parameter (i.e. the amount of time needed to complete a reaching task). As such, the hypotheses for this study were:

H_1 : Participants will effectively decrease their MTs while receiving feedback cues from a robotic playmate while interacting with our *Super Pop VRTM* system.

H_2 : Participants' MTs will still be reduced even after the feedback cues provided by the robotic playmate are withdrawn.

7.3.2 Experimental Design

Seven children with cerebral palsy (CP) and ten typically developing (TD) children were recruited to interact with our *Super Pop VRTM* system (Figure ??). There were four females and three males in the CP group (mean age = 9.86 years, standard deviation = 1.35 years), and seven females and three males in the TD group (mean age = 9.60 years, standard deviation = 1.26 years). The children's parents signed the IRB (Institutional Review Board) approved consent form allowing them to engage in the testing sessions.

To validate the study's hypotheses and determine if and how participants modified their kinematic behavior, their performances before, during, and after a training session with our *Super Pop VRTM* system were compared to each other. The setup of the game settings is described in Appendix ?. Participants' performances were evaluated with respect to the 90° reaching task described in Figure ?. Each participant interacted with the system for three rounds and performed the reaching task

20-30 times for each round. For the first and third rounds, participants performed the reaching task without receiving feedback. For the second round (training session), participants received instructional and corrective feedback from the robotic playmate (refer to Section ?? for more details on Darwin). The architecture of the testing sessions is similar to the one described in Figure ?. However, instead of evaluating if the participant reached a goal, the testing session stops when the participant completed the required amount of reaching tasks. Moreover, instead of following the feedback cues described in Table ?, Darwin would say *“Keep up the good work. Move a little faster.”* if the participant’s MT was greater than the targeted threshold (TH). Similarly, if the participant’s MT was less than or equal to the TH, Darwin would say *“Fantastic. Let’s move at the exact same speed.”*. For each participant, the movement time threshold (MT_{TH}) was defined as 80% of his/her baseline MT (i.e. the participant’s average MT from round 1). In this manner, we prompt the participant to continually move at a pace faster than their natural speed.

During game play, the system recorded the MT taken by participants to complete each reaching task. For each round, let random variable X be a participant’s MT after completing a reaching task. With no prior information about the underlying distribution of a participant’s MTs, we adhere to the Central Limit Theorem (i.e. the sum of many random variables will have, approximately, a normal distribution), and assume $X \sim N(\mu, \sigma)$ to be normally distributed, where μ and σ are the mean and standard deviation, respectively, of the participant’s MTs for the corresponding round. For each round, we compute the probability that a participant will complete a reaching task with a MT less than or equal to his/her corresponding MT_{TH} (i.e. $F_X(x) = P(X \leq x)$, where x is the participant’s MT_{TH}). By definition, this is the CDF (cumulative distribution function) of X , which is given by (??), given our assumption that X is normally distributed.

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt \quad (40)$$

7.3.3 Results

Based on the collected MT data for each round, the probabilities that a participant will complete a reaching task with a MT less than or equal to his/her corresponding MT_{TH} are shown in Table ???. Moreover, we aggregate the results and organize them by group: typically developing (TD) kids and kids who have cerebral palsy (CP). For each group, boxplots showing how the probabilities are distributed with respect to each round are shown in Figure ??. As an example, boxplots showing how the MTs of one participant, from each group, are distributed with respect to each round are shown in Figure ??.

Table 23: Probabilities that the participants' MTs will be less than or equal to their corresponding THs for each phase.

Typically Developing Children				Children with Cerebral Palsy			
Participants	R1	R2	R3	Participants	R1	R2	R3
1	0.232	0.289	0.161	1	0.240	0.312	0.290
2	0.224	0.664	0.446	2	0.253	0.319	0.341
3	0.224	0.580	0.575	3	0.213	0.245	0.261
4	0.314	0.574	0.999	4	0.122	0.594	0.212
5	0.257	0.348	0.385	5	0.214	0.631	0.432
6	0.243	0.372	0.369	6	0.317	0.590	0.503
7	0.129	0.423	0.214	7	0.252	0.452	0.303
8	0.344	0.455	0.797	-	-	-	-
9	0.182	0.337	0.153	-	-	-	-
10	0.219	0.359	0.395	-	-	-	-
AVG	0.237	0.436	0.450	AVG	0.230	0.449	0.335
STD	0.061	0.129	0.275	STD	0.059	0.159	0.101

* R1, R2, and R3 make reference to rounds 1, 2, and 3 respectively.

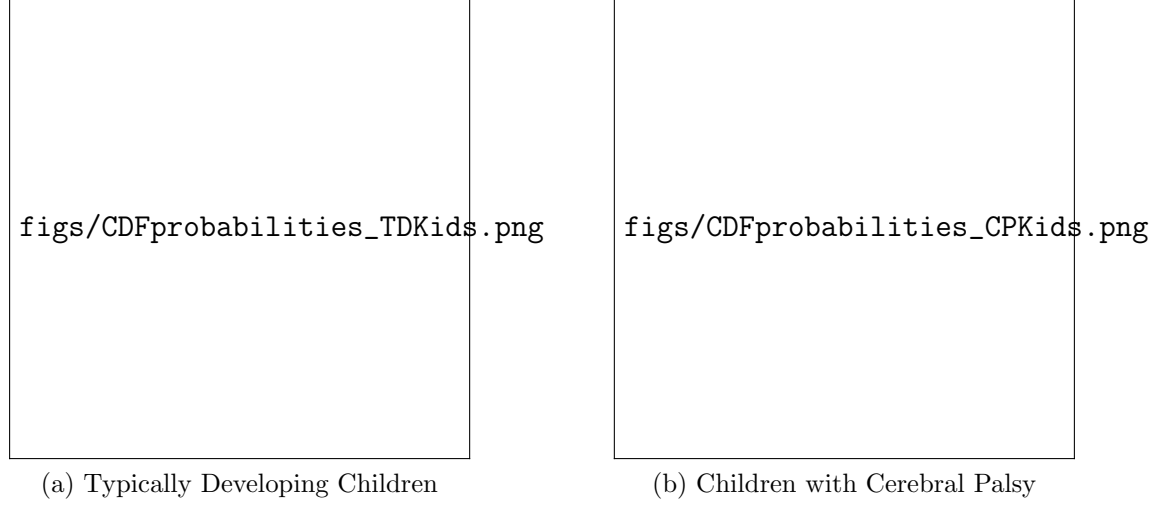


Figure 34: Boxplots with the distributions, with respect to each round, of the probabilities that a participant will complete a reaching task with a MT less than or equal to his/her corresponding MT_{TH} .

7.3.4 Discussion and Conclusions

The averages in Table ?? and the boxplots in Figure ?? show that, for both TD children and children who have CP, the probability for a participant to complete a reaching task with a MT less than or equal to the corresponding MT_{TH} is greater in rounds 2 and 3 than in round 1. As such, while receiving corrective feedback during game play in round 2, participants are more likely to complete a reaching task with a lower MT than their baseline from round 1, thus validating our first hypothesis (H_1). Similarly, after removing the feedback cues in round 3, although the probability for completing a reaching task is lower than the training session in round 2, it is still higher than the baseline from round 1. This validates our second hypothesis (H_2) since participants are still more likely to complete a reaching task with a lesser MT even after withdrawing the feedback cues.

We conducted four separate unpaired two-tailed t -tests at a 95% confidence level to determine if the probabilities in Table ?? and Figure ?? are statistically different between rounds. The null-hypothesis is the same for the four tests: the probabilities

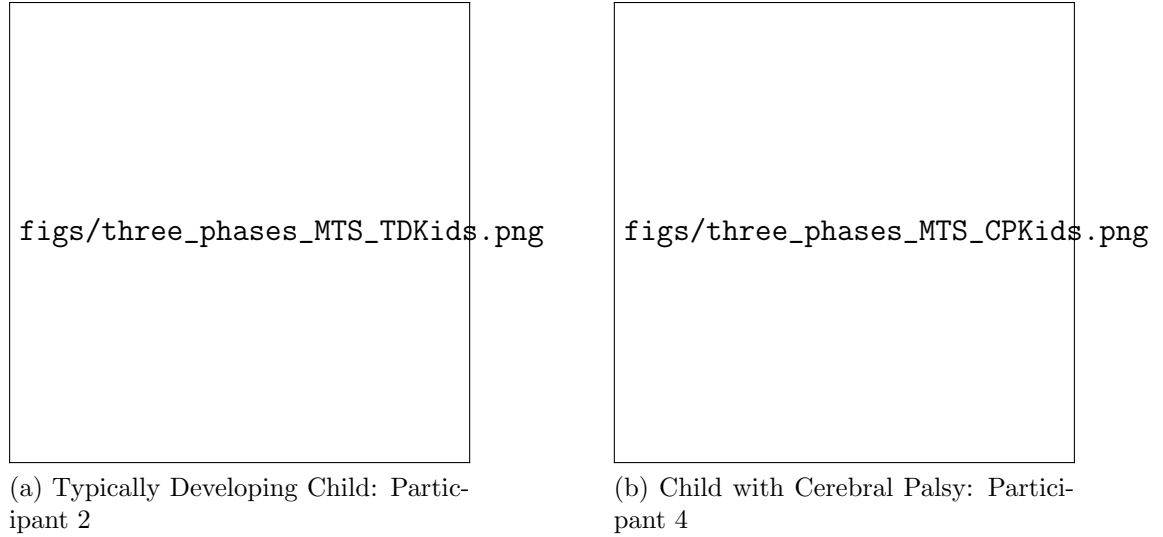


Figure 35: Boxplots with the MTs for one participant organized with respect to each round.

between two rounds come from distributions with equal means. The resulting p -values are shown in Table ?? . Seeing as $p < 0.05$ for all tests, we reject all null-hypotheses and conclude that there is a statistical difference between the rounds for both groups. Namely, between round 1 and round 2 showing that participants are more likely to decrease their MTs while receiving corrective feedback during game play (hypothesis H_1), and between round 1 and round 3 showing that participants still have a higher probability to decrease their MTs even after withdrawing the corrective feedback (hypothesis H_2).

Table 24: p -values at a 95% confidence level for testing statistical difference in probabilities from Table ?? between rounds.

	R1 : R2	R1 : R3
Typically Developing Children	$\ll 0.05$	0.028
Children with Cerebral Palsy	0.005	0.036

*‘R1 : R2’ and ‘R1 : R3’ refer to the t -tests between Round 1 and Round 2, and Round 1 and Round 3 respectively.

These trends are observed when analyzing each participant’s MT distribution individually. For example, the MTs of Participant 2 (from the TD group) and Participant

4 (from the group with CP) decrease in round 2 when the participants receive the corrective cues during game play (Figure ??). Moreover, their MTs increase in round 3 but are still less than their baseline from round 1. These observations further support the study's hypotheses. Thus, by prompting participants to complete the reaching tasks at a pace faster than their baseline (i.e. to complete the reaching tasks with a MT less than 80% of their baseline), they were able to improve their performance while receiving feedback and maintain their improvement after removing the feedback cues.

There are a few limitations with this study. The first limitation is that participants' performance was evaluated with respect only to the movement time parameter. Future studies should be conducted to evaluate if the hypotheses still hold with respect to other kinematic parameters and/or combinations of parameters. Another limitation is in the number of reaching tasks completed by each participant. Results are obtained based on the assumption that the participants' MTs in a given round are normally distributed. We would have better information about the underlying distribution if participants completed more reaching tasks for each round. Finally, all rounds were performed one after another on the same day. As such, longitudinal studies should be conducted to determine whether participants can maintain their improvements over longer periods of time after interacting with our system.

7.4 *Summary*

It is our ultimate goal to develop a system that can provide targeted feedback for any and all kinematic parameters of interest for individuals with some form of motor skills disorders. Such a system, designed for an in-home environment, would increase the frequency at which users receive feedback regarding their kinematic performance thus increasing the efficacy of their intervention protocols. As such, this chapter focused on evaluating if our system has the capability of prompting users to modify

and adapt their kinematic behavior via feedback. We conducted two pilot studies in which our system provided different types of corrective feedback and evaluated the behavior response of the participants. Both studies evaluated the system as described in Figure ??, used the humanoid robot DARwIn-OP as the robotic playmate that provides corrective feedback (Section ??), and adhere to the findings of our independent study that showed that a combination of verbal and nonverbal cues is the most effect manner for providing corrective feedback (Section ??).

The first pilot study focused on determining if individuals that interacted with our system can reach a specific performance goal (Section ??). Not only do results show that adults interacting with the version of the system that provides the corrective feedback via a virtual agent can reach their individualized movement time (MT) references, results also show that receiving corrective feedback from an embodied robotic playmate will help individuals reach their MT references faster (i.e. in a less amount of trials) than receiving feedback from a virtual agent. Results from the second round of experiments show that typically developing children that interact with the version of our system that provides corrective feedback via the robotic playmate can reach their individualized MT references as well. The results from this study suggest that our system can accomplish one of the main objectives of physical therapy with a human therapist: to reach a performance goal over an extended period of time.

The second pilot study focused on evaluating if participants could improve their performance during a training session with our system (Section ??). Results show that both typically developing children and children who have cerebral palsy can, not only improve their kinematic performance (i.e. decrease their MTs for this study) during a training session with our system, but also maintain their improvement after the corrective cues are removed. The results from this study suggest that users can improve their kinematic performance by interacting with our *Super Pop VRTM* system, as well as maintain it afterwards.

CHAPTER VIII

CONCLUSIONS AND FUTURE WORK

The main focus of this dissertation research is to increase the rate of improvement in individuals with some form of motor skills disorder through interaction with an engaging serious game that not only assesses the user's kinematic performance, but also continuously provides targeted corrective feedback. The benefits of standard physical therapy protocols have been well documented and are summarized in our literature review in Chapter ???. To address the problem of non-compliance to perform the recommended in-home therapy exercises due to lack of motivation, work has been done in developing serious gaming systems that increase user motivation and engagement (Section ??). While there have been advances in this field, our literature surveys reveals that existing systems do not employ an assessment methodology that objectively and quantifiably evaluates the user's kinematic performance (Section ??), nor do they provide the targeted corrective feedback necessary for users to judge and improve their performance (Section ??).

As such, this dissertation research aims to fill in the gap by designing, developing, and validating a more robust system that can provide the necessary targeted corrective feedback as a function of the objective assessment of the user's kinematic performance (Figure ??). This dissertation describes the design process, the algorithms developed, the models constructed, and the the user studies conducted to validate the overall system and our claims. The following sections summarize this research's key contributions, the publications that resulted from our work, and our recommendations for future work.

8.1 Contributions

8.1.1 Virtual Reality Serious Game for Rehabilitation

We designed and developed a serious gaming system that can be used as part of the intervention protocols for individuals who have some form of motor skills disorder. Chapter ?? describes the features and functionality of our system: the *Super Pop VRTM* game. Because of its portability, the game uses the KinectTM camera from Microsoft as its sensing methods such that it can be used outside of the clinical setting, for example, in the home setting. Moreover, the game allows for the individualization of therapy protocols by allowing the therapist or clinician to select the combination of game settings that best addresses the user's needs. In our research, for example, we employed an algorithm that selects an appropriate evaluation trajectory as a function of the user's upper-body dimensions and functional capabilities (Algorithm ??).

Section ?? describes the engagement study we conducted to determine the level of engagement and self-reported intrinsic motivation users experience when interacting with the *Super Pop VRTM* game. The study concluded that, in general, the *Super Pop VRTM* game promotes a relatively relaxing environment where users can perform their therapy exercises with minimal pressure or tension while having a sense of competence in their actions. Results suggest that our system can address the problem of non-compliance when it comes to performing the recommended in-home therapy exercises.

8.1.2 Real-time Generation of Baseline Movements

Instead of going through the time-consuming and potentially tedious process of collecting human data to construct the baseline models for the user's upper-body assessment, we apply a kinematic model of the human arm to generate baseline trajectories for a given reaching task in real-time (Section ??). It has been shown that kinematic measures during a reaching task are correlated to functional measures of

upper-extremity function, and that kinematic measures can be used to assess performance. As such, we defined a set of kinematic parameters that mathematically describe an individual's upper body movements, which are used as the basis for objectively and quantifiably assessing users' upper-body movements (Section ??). As part of this contribution, we also conducted two separate user studies to examine and validate the performance of the kinematic model's performance relative to the defined kinematic parameters.

The first pilot study examines the kinematic model's capability of yielding similar values as those produced by the user relative to the elbow and shoulder (ROM) parameters (Section ??). We analyzed the percent error differences between the participants' outcome metrics and the baseline values generated by the kinematic model. The study concluded that, since the results fall in-line with error ranges obtained from human-data collection approaches, the proposed kinematic model can accurately represent the movement kinematics of the human arm relative to the elbow and shoulder ROM parameters.

The second pilot study validates the baselines generated by our system with respect to all seven kinematic parameters (Section ??). For the kinematic parameters based on our kinematic model, results show that the differences between the sample means of the data collected from humans and the data computed by our kinematic model are small enough such that we consider the two baselines statistically similar to each other with respect to the deviation from line, and elbow and shoulder range of motion parameters. For the kinematic parameters that are based on human models, results show that, a baseline constructed with data computed from external human models would yield a more accurate description of how humans complete movement tasks when compared to a baseline constructed from the data directly collected from participants because of its high variability.

8.1.3 Feasibility of the Super Pop System as an Evaluation Tool

We validated the feasibility of using the *Super Pop VRTM* game as a reliable and accurate evaluation tool to measure individuals' reaching kinematics by quantifying the accuracy of the sensing methods, examining the system's feasibility of being used as an evaluation tool, and examining the test-retest reliability of the overall system. The pilot study described in Section ?? evaluated the KinectTM camera's performance with respect to the OptiTrack, a current state-of-the-art motion capture system. Results show that 1) both motion capture systems yield similar outcome measures when evaluating users' reaching kinematics with the *Super Pop VRTM* game, and 2) the trajectories generated by both tracking systems are spatially similar to each other. These results support the validity of using the Kinect as the motion capture system for home-based rehabilitation purposes.

The pilot study described in Section ?? focused on: 1) validating the *Super Pop VRTM* game as a feasible system for the in-home setting for documenting arm function improvement in children who have cerebral palsy (CP) relative to their reaching kinematics, and 2) examining the analysis using reaching kinematics as measured by the *Super Pop VRTM* game with respect to two standardized clinical assessment methodologies. The study concludes that the *Super Pop VRTM* game is feasible for documenting improvement in children's reaching kinematics seeing as: the system reliably kept track of the participants' reaching kinematics, the system was easy to assemble and implement, and the participants enjoyed playing the game without noticing that their movements were quantitatively measured. In addition, results also showed that there was no statistically significant difference between mean values of the kinematic parameters in children with CP and those with typical development, after an 8-week home-based virtual reality intervention for improving children's arm function. Unfortunately we were not able to confirm a correlation between the reaching kinematics analysis measured by the *Super Pop VRTM* game and the analysis

made by two standardized clinical assessment methodologies given their subjectivity.

Section ?? discussed the testing sessions conducted to determine the *Super Pop VRTM* game’s test-retest reliability within and between days. After analyzing the intraclass correlation coefficients (ICC) computed relative to a set of kinematic parameters, the study concludes that the game has good to excellent reliability between sessions and between days for most parameters. These results serve as an extension of the study described in Section ??, further confirming the potential for using the *Super Pop VRTM* game as a reliable evaluation tool to measure individuals’ reaching kinematics.

8.1.4 Movement Classification and Baseline Selection

Previous studies have shown that, as part of physiotherapy protocols, external feedback of performance is essential to motor learning. Moreover, it has also been shown that external feedback is most efficient when provided relative to the user’s abilities. As such, for our system to have the capability to efficiently provide targeted corrective feedback, it first needs the capability to identify the user’s current kinematic level. We achieved this with our fourth contribution by training a pattern recognition classification model that identifies the user’s kinematic class. This refers to classifying the user’s upper-body reaching kinematics as a function of his/her abilities. Continuous identification of the user’s kinematic class allows the system to autonomously select the most appropriate baseline model such that the feedback is targeted relative to his/her kinematic performance at any given point in time. Chapter ?? describes the experiments conducted to compare the performances of different pattern recognition classification methodologies and select the one that best separates the classes of interest. For each new reaching task, the system can then apply the final trained model to identify the user’s current kinematic class, and use that information to select the baseline model that best fits the user’s ability.

8.1.5 Kinematic Behavior Adaptation

With all previous contributions leading to a functional system, the next step was to evaluate its efficacy in prompting users to adapt and modify their kinematic behavior as a function of the provided corrective feedback. We conducted two users studies to determine if, via interactions with our system, individuals can: 1) modify their kinematic behavior such that they can reach an individualized performance goal (Section ??), and 2) improve their kinematic performance while continuously receiving corrective feedback and then maintaining said performance after the feedback is removed (Section ??).

The first pilot study focused on determining if individuals that interacted with our system can reach a specific performance goal (Section ??). Not only do results show that adult participants interacting with the version of the system that provides the corrective feedback via a virtual agent can reach their individualized movement time (MT) references, results also show that receiving corrective feedback from an embodied robotic playmate will help individuals reach their MT references faster (i.e. in a lower amount of trials) than receiving feedback from a virtual agent. Results from the second round of experiments show that typically developing children that interact with the version of our system that provides corrective feedback via the robotic playmate can reach their individualized MT references as well. The results from this study suggest that our system can accomplish one of the main objectives of physical therapy with a human therapist: to reach a performance goal over an extended period of time.

The second pilot study focused on evaluating if participants could improve their performance during a training session with our system (Section ??). Results show that both typically developing children and children who have cerebral palsy can, not only improve their kinematic performance (i.e. decrease their MTs for this study) during a training session with our system, but also maintain their improvement after

the corrective cues are removed. The results from this study suggest that users can improve their kinematic performance by interacting with our *Super Pop VRTM* system, as well as maintain it afterwards.

8.2 Publications

The following refereed publications were derived from this dissertation:

8.2.1 In Preparation

1. Y.P. Chen, **S. García-Vergara**, A.M. Howard, “Examining the Effect of Feedback from a Humanoid Robot on Reaching Kinematics in Children with Cerebral Palsy,” *American Physical Therapy Association NEXT Conference*, 2016.

8.2.2 Book Chapter

1. **S. García-Vergara**, L. Brown, H.W. Park, and A.M. Howard, “Engaging children in play therapy: The coupling of virtual reality games with social robotics,” *Technologies of Inclusive Well-Being*, Springer Berlin Heidelberg, pp. 139-163, 2014.

8.2.3 Journals

1. Y.P. Chen, **S. García-Vergara**, and A.M. Howard, “Effect of a Home-Based Virtual Reality Intervention for Children with Cerebral Palsy using *Super Pop VRTM* Evaluation Metrics: A Feasibility Study,” *Rehabilitation Research and Practice*, 2015.

8.2.4 Refereed Conference Publications

1. **S. García-Vergara**, P. Robinette, Y.P. Chen, and A.M. Howard, “Validation of a Physical Rehabilitation Game using Markerless versus Marker-based Motion Capture Systems,” *IEEE EMBS Conference*.

2. **S. García-Vergara**, L. Brown, Y.P. Chen, and A.M. Howard, “Increasing the Efficacy of Rehabilitation Protocols for Children via a Robotic Playmate Providing Real-time Corrective Feedback,” *25th IEEE RoMan Conference*, pp. 700-705, 2016.
3. L. Brown, **S. García-Vergara**, and A.M. Howard, “Evaluating the Effect of Robot Feedback on Motor Skill Performance in Therapy Games,” *IEEE Conference on Systems, Man, and Cybernetics (SMC)*, pp. 1060-1065, 2015.
4. **S. García-Vergara**, H. Li, and A.M. Howard, “Increasing *Super Pop VRTM* Users’ Intrinsic Motivation by Improving the Game’s Aesthetics,” *International Conference on Universal Access in Human-Computer Interaction*, pp. 432-441, 2015.
5. **S. García-Vergara**, M.M. Serrano, Y.P. Chen, and A.M. Howard, “Developing a Baseline for Upper-body Motor Skill Assessment Using a Robotic Kinematic Model,” *IEEE RoMan Conference*, pp. 911-916, 2014.
6. **S. García-Vergara**, and A.M. Howard, “Three-dimensional Fitts Law Model used to Predict Movement Time in Serious Games for Rehabilitation,” *International Conference on Virtual, Augmented and Mixed Reality*, pp. 287-297, 2014.
7. **S. García-Vergara**, Y.P. Chen, and A.M. Howard, “*Super Pop VRTM* : an Adaptable Virtual Reality Game for Upper-Body Rehabilitation,” *International Conference on Human-Computer Interaction*, pp. 40-49, 2013.

8.3 Recommendations for Future Work

8.3.1 Movement Classification and Baseline Model Construction

Chapter ?? describes the importance of selecting the most appropriate baseline model for an accurate assessment of the user’s upper-body movements. As part of this

dissertation research we only constructed one baseline model (see Section ?? for details on the model constructed for the movement time kinematic parameter). The full functionality of our system depends on its ability to select from a wide range of baseline models such that it can accommodate as many user demographics as possible. Additional research should be conducted to develop these models, or even better, develop an approach that can autonomously construct baseline models as a function of key descriptive features that best describe a given user demographic.

Chapter ?? also describes the approaches we experimented with to train our pattern recognition classification model. The results shown are relative to two general classes: upper-body movement profiles with characteristics of children who have cerebral palsy and movement profiles without said characteristics. The ultimate goal is for the system to be able to identify the user’s kinematic level with as much precision as possible. To accomplish this, additional research should be conducted such that the final classification model can account for more than these two general classes. That is to say, the model should be able to distinguish between enough classes that would could cover the majority of the patient demographics.

8.3.2 Longitudinal Clinical Studies

In general, **longitudinal studies** are defined as studies in which the performance of the participants is evaluated on two or more occasions [?]. Previous studies have determined longitudinal studies to be one of the principal research strategies employed in medical research [?, ?]. Although we evaluate the participants’ performance on more than two occasions for the user studies we conducted as part of this dissertation research, the duration of the studies was, on average, approximately 30-90 minutes. This is not enough time to observe the long-term effects of interacting with our system. As such, to further support our claims, future studies should be of longer

durations such that we can attribute any improvements in the participants' arm function to their interactions with the game and not to potentially random and temporary changes.

We would ultimately like to obtain all the necessary evidence to show that our *Super Pop VRTM* game (Figure ??) has the capability of identifying the user's kinematic level, and of continuously providing targeted corrective feedback relative to the autonomous selection of the baseline model that is closest to the user's abilities. More specifically, we would like to show that individuals can effectively improve their arm function by efficiently being prompted to adapt and modify their kinematic behavior based on the corrective feedback they receive relative to their kinematic performance. It's important to note that this dissertation makes no comparison between the amount of arm function improvement an individual experiences by participating in typical physiotherapy protocols versus by interacting with the *Super Pop VRTM* system. Thus, longitudinal clinical studies might aid in validating our claim that continuous interaction with our system will increase the user's rate of improvement in arm function.

8.3.3 Additional Clinical Studies

As part of this dissertation research, we conducted two user studies to examine our system's capability of prompting users to adapt and modify their kinematic behavior relative to the provided corrective feedback. However, the studies described in Chapter ?? assume a constant baseline model and do not directly integrate the movement classification phase (Figure ??). To validate the full functionality of our *Super Pop VRTM* system, additional user studies will need to be conducted using the complete integrated system architecture as depicted in Figure ??.

APPENDIX A

SUPER POP VR™'S MANUAL AND TROUBLESHOOTING

appendixdocs/SuperPopManual.pdf

appendixdocs/SuperPopManual.pdf

appendixdocs/SuperPopManual.pdf

appendixdocs/SuperPopManual.pdf

appendixdocs/SuperPopManual.pdf

appendixdocs/SuperPopManual.pdf

appendixdocs/SuperPopManual.pdf

appendixdocs/SuperPopManual.pdf

APPENDIX B

IMI SURVEY FOR ENGAGEMENT STUDY

appendixdocs/IMISurvey.pdf

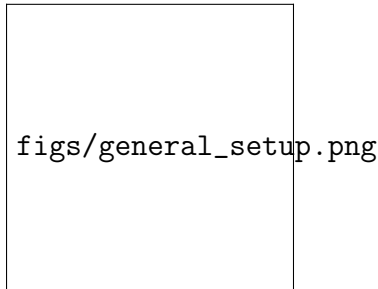
appendixdocs/IMISurvey.pdf

APPENDIX C

PROTOCOL FOR REGULAR SUPER POP VRTM EXPERIMENTS

A. General Description

This is the main protocol followed during testing sessions of the *Super Pop VRTM* game. Before each testing session, the game was setup using the manual in Appendix ???. For all testing sessions, the 3D coordinates of the participants' upper-body joints were collected and stored during the process of completing the 90° reaching task described in Figure ??. All participants interacted with the system in their homes. Different testing sessions have different time durations depending on the purpose. The environment settings were maintained as a constant in order to maintain consistency. For all sessions, the game interface was projected onto a large screen via a projector connected to a PC laptop. Some examples are shown below in the figures below:



B. Measurements

Measurements	Measures
Chair Height	41cm
Distance between user's chair and Kinect camera	190cm
Distance between projector and projecting screen	170cm

C. Script and Administrator Instructions

- “Hello. My name is Sergio. Thank you for taking time to participate in this study. Today you will play a game that keeps track of your upper-body movements. The Kinect camera will track your movements and map them into the screen.”
- *Press the ‘play’ button and verify that the virtual markers are following the participant’s hands and head.*
- “When you move your arms, can you see the blue markers that follow your hand?” [*if yes*], “During the game you will move your arms and use these markers to pop the bubbles that will appear on screen.” [*if no, restart game until participant can see blue markers follow their hands.*]
- “The goal is to pop as many bubbles as you can to get as many points as you can. The yellow bubbles are worth 5 points, the green bubbles are worth 10 points, and the red bubbles are worth -5 points. Try to get as many points as possible.”
- “You will play several different games, but let’s calibrate the system first. Please put both arms as high as you can.” *Open the Bubble Appearance Region GUI.* “You can put your arms down thank you.”
- *Place the super bubbles with a separation of 90° between all of them. Make sure that each super bubble trajectory will appear every 5 seconds, that each super bubble will last 10 seconds on screen, that there are three super bubbles, and that we’re assessing the correct arm (dominant/preferred or non-dominant/affected). Close the interface.*
- *Select the corresponding username. Start the game.*