From Head to Toe:

Body Movement for Human-Computer Interaction

Eduardo Velloso

Graduate College

School of Computing and Communications

Lancaster University

This dissertation is submitted for the degree of Doctor of Philosophy

April, 2015

To my grandfather Dr. Dirceu de Alencar Velloso (1931-2005),
Professor of Civil Engineering, Emeritus,
the source of inspiration for all my academic endeavours

"Ó vida futura! Nós te criaremos."

Carlos Drummond the Andrade

DECLARATION

This thesis is the result of my own original research. It has not been previously submitted, in part or whole, to any university of institution for any degree, diploma, or other qualification. The work described in Chapters 2 and 3 were conducted under the supervision of Hans Gellersen and Andreas Bulling, and the work described in the remaining chapters also counted with the additional supervision of Jason Alexander. Dominik Schmidt helped with the survey presented in Chapter 4, and Adalberto Simeone conducted the study in Section 6.2. I was responsible for the concept, supervision, management, and game design aspects of the arcade machine. Carl Oechsner, Katharina Sachmann and Markus Wirth helped with designing and building the machine, as well as implementing two of the games.

Signed:_____

Date: 28/04/2015

Eduardo Velloso

April, 2015

Lancaster, UK

ABSTRACT

Our bodies are the medium through which we experience the world around us, so human-computer interaction can highly benefit from the richness of body movements and postures as an input modality. In recent years, the widespread availability of inertial measurement units and depth sensors led to the development of a plethora of applications for the body in human-computer interaction. However, the main focus of these works has been on using the upper body for explicit input. This thesis investigates the research space of full-body human-computer interaction through three propositions.

The first proposition is that there is more to be inferred by natural users' movements and postures, such as the quality of activities and psychological states. We develop this proposition in two domains. First, we explore how to support users in performing weight lifting activities. We propose a system that classifies different ways of performing the same activity; an object-oriented model-based framework for formally specifying activities; and a system that automatically extracts an activity model by demonstration. Second, we explore how to automatically capture nonverbal cues for affective computing. We developed a system that annotates motion and gaze data according to the Body Action and Posture coding system. We show that quality analysis can add another layer of information to activity recognition, and that systems that support the communication of quality information should strive to support how we implicitly communicate movement through nonverbal communication. Further, we argue that working at a higher level of abstraction, affect recognition systems can more directly translate findings from other areas into their algorithms, but also contribute new knowledge to these fields.

The second proposition is that the lower limbs can provide an effective means of interacting with computers beyond assistive technology To address the problem of the dispersed literature on the topic, we conducted a comprehensive survey on the lower body in HCI, under the lenses of users, systems and interactions. To address the lack of a fundamental understanding of foot-based interactions, we conducted a series of studies that quantitatively characterises several aspects of foot-based interaction, including Fitts's Law performance models, the effects of movement direction, foot dominance and visual feedback, and the overhead incurred by using the feet together with the hand. To enable all these studies, we developed a foot tracker based on a Kinect mounted under the desk. We show that the lower body can be used as a valuable complementary modality for computing input.

Our third proposition is that by treating body movements as multiple modalities, rather than a single one, we can enable novel user experiences. We develop this proposition in the domain of 3D user interfaces, as it requires input with multiple degrees of freedom and offers a rich set of complex tasks. We propose an approach for tracking the whole body up close, by splitting the sensing of different body parts across multiple sensors. Our setup allows tracking gaze, head, mid-air gestures, multi-touch gestures, and foot movements. We investigate specific applications for multimodal combinations in the domain of 3DUI, specifically how gaze and mid-air gestures can be combined to improve selection and manipulation tasks; how the feet can support the canonical 3DUI tasks; and how a multimodal sensing platform can inspire new 3D game mechanics. We show that the combination of multiple modalities can lead to enhanced task performance, that offloading certain tasks to alternative modalities not only frees the hands, but also allows simultaneous control of multiple degrees of freedom, and that by sensing different modalities separately, we achieve a more detailed and precise full body tracking.

ACKNOWLEDGEMENTS

This thesis would not have been possible without the always excellent advice and guidance from my three supervisors, Prof. Hans Gellersen, Dr. Andreas Bulling, and Dr. Jason Alexander. Throughout my PhD, they kept me inspired and motivated and taught me countless lessons that I will take with me for the rest of my life.

I would also like to thank Dr. Hugo Fuks for inspiring me to follow a career in research, and Dr. Alessandro Garcia for making the introductions that led me to Lancaster.

I am immensely grateful for the contributions of the incredible researchers with whom I collaborated as co-author in the papers that I wrote during my PhD, including Hans Gellersen, Andreas Bulling, Jason Alexander, Hugo Fuks, Wallace Ugulino, Adalberto Simeone, Jayson Turner, Faisal Taher, Dominik Schmidt, and Augusto Esteves.

No man is an island, and I was lucky to be a member of a fantastic research group who consistently kept the bar high and motivated me to always strive for excellence in my research: Jayson Turner, Adalberto Simeone, Ken Pfeuffer, Christian Weichel, Augusto Esteves, Mélodie Vidal, Yanxia Zhang, Carl Fischer, Dominik Schmidt, Ming Ki Chong, Matt Oppenheim, John Hardy, Faisal Taher, Pierre Weill-Tessier.

I could not have gotten here without the loving support from my family and friends, especially my parents Bia and Milton, my sister Ana Clara, and my lifelong friends Manos.

Finally, I am eternally grateful for the love and care of my Kasinha. During these years you have been an amazing source of strength for me, always pushing me to become the best I can be, being there for me when I most needed and helping me with all aspects of my work.

CONTENTS

LIST OF TABLES

List of Figures

RELATED PUBLICATIONS

Some of the work presented in this thesis was published in the following papers:

- [276] Velloso, E., Bulling, A. and Gellersen, H. 2013. AutoBAP: Automatic coding of body action and posture units from wearable sensors. Affective Computing and Intelligent Interaction (ACII), 2013 Humaine Association Conference on (2013), 135–140.
- [277] Velloso, E., Bulling, A. and Gellersen, H. 2013. MotionMA: Motion Modelling and Analysis by Demonstration. Proc. of the 31st SIGCHI International Conference on Human Factors in Computing Systems (2013).
- [275] Velloso, E., Bulling, A. and Gellersen, H. 2011. Towards qualitative assessment of weight lifting exercises using body-worn sensors. Proceedings of the 13th international conference on Ubiquitous computing (2011), 587–588.
- [274] Velloso, E., Bulling, A., Gellersen, H., Ugulino, W. and Fuks, H. 2013. Qualitative activity recognition of weight lifting exercises. Proceedings of the 4th Augmented Human International Conference (2013), 116–123.
- [242] Simeone, A. L., Velloso, E., Alexander, J., and Gellersen, H. 2014. Feet movement in desktop 3D interaction. Proceedings of the 2014 IEEE Symposium on 3D User Interfaces (2014), 71-74.
- [278] Velloso, E., Cardador, D., Vega, K., Ugulino, W., Bulling, A., Gellersen, H., and Fuks, H. 2011. The Web of Things as an Infrastructure for Improving Users' Health and Wellbeing. Proceedings of II Symposium of the Brazilian Institute for Web Science (2011).
- [273] Velloso, E., Alexander, J., Bulling, A., and Gellersen, H. 2015. Interactions under the Desk: A Characterisation of Foot Movements for Input in a Seated Position. Proceedings of the IFIP International Conference on Human-Computer Interaction (2015)
- [279] Velloso, E., Turner, J., Alexander, J., Bulling, A., and Gellersen, H. 2015. An Empirical Investigation of Gaze Selection in Mid-Air Gestural 3D Manipulation. Proceedings of the IFIP International Conference on Human-Computer Interaction (2015)

  Velloso, E., Schmidt, D., Alexander, J., Gellersen, H., and Bulling, A. 2015. The Feet in HCI: A Survey of Foot-Based Interaction. ACM Computing Surveys, 48-2, pp. 21:1—21:35 (2015)
- Velloso, E., Oechsner, C., Sachmann, K., Wirth, M. and Gellersen, H. 2015. Arcade+: A Platform for Public Deployment and Evaluation of Multi-Modal Games. Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play, 271-275 (2015)

1 INTRODUCTION

"The human body is the best picture of the human soul."

Ludwig Wittgenstein

Since the inception of the field, research on Human-Computer Interaction has strived to find more efficient and natural ways of interacting with computing systems. With the advent of inexpensive inertial measurement units (IMU) and consumer-ready depth cameras (e.g. Kinect), Bodily Interaction has surfaced as a natural, intuitive and highly expressive way of providing input to computers. These capabilities are now present in a multitude of devices we have at home, including mobile phones, watches and video game consoles. By allowing users to freely use their bodies as input, these systems enable interactions using gestures, postures and actions that more closely resemble how we manipulate the physical world around us. In this thesis, we identified three directions in which the research space of bodily interaction can be expanded: implicit interaction, lower body interaction, and multimodal interaction.

One topic in Bodily Interaction that has been widely explored is that of body movements to issue commands and manipulate controls—what we refer to as explicit input. Implicit interaction represents the other side of explicit input. In the context of bodily interaction, this usually leads to systems that observe users' natural movements and postures and use this information to make conclusions about the activities and the context of use. Open challenges for implicit interaction include inferring not only *which* activity is being performed, but *how* it is being performed; inferring deeper insights about users, such as their psychological states; and what to do with this information after it has been inferred.

A considerable amount of work has explored applications for bodily interaction, including work on mid-air gestures, multi-touch surfaces and tangible computing. However, most of this work concentrated on using the upper body—fingers, hands, arms, torso, head, and eyes—for these tasks. Research on lower limbs usually focuses on accessible input (e.g. foot mice) or monitoring (e.g. smart trainers). Therefore, whereas we have a thorough understanding of the capabilities of the upper body for bodily interaction, the same cannot be said about the legs and feet.

Multimodal systems are those that take input or provide output through multiple channels. As the body can offer multiple degrees of freedom of input, different body parts can be interpreted as separate input modalities. The wide availability of inexpensive sensors created the opportunity for incorporating multimodality in diverse application scenarios. The body itself can be seen a source of multiple input channels for systems, making full body interactions inherently multimodal.

We propose to investigate the research space of bodily interaction by tackling open research questions in these three directions: implicit interaction, lower limbs and multimodal interaction. This thesis is therefore guided by three propositions. The first proposition is that **there is more to be inferred by natural users' movements and postures, such as the quality of activities and psychological states**. Our second proposition is that **the lower limbs can provide an effective means of interacting with computers beyond assistive technologies**. Our third and final proposition is that **by treating body movements as multiple modalities, rather than a single one, we can enable novel user experiences**.

Based on these propositions, we derived a series of research questions that we explore in the following chapters. The first question is **how do we specify movements for quality analysis?** We propose and evaluate several approaches for this problem in the domains of weight lifting and affective computing. The second question is **how can the feet be used for direct input to interactive systems?** To answer this, we not only look at the past research on foot-based interaction, but also conduct experiments to characterise foot-based input and propose novel interaction techniques. The third question is **how can combinations of input modalities support complex interactive tasks?** We explore this question in the domain of 3D user interfaces by investigating several modalities, including mid-air gestures, gaze, feet, and touch.

1.1 Implicit Interaction

Our first proposition relates to implicit interaction. Schmidt defines implicit human-computer interaction as "an action performed by the user that is not primarily aimed to interact with a computerised system but which such a system understands as input" [232]. Several categories of systems take advantage of implicit interaction, including context-aware systems, human activity recognition systems and affective computing systems. Because of the high communicative power of body movements and postures, we are particularly interested in systems that capture implicit information from the body.

Context-aware systems are those that attempt to characterise the situation of a person, place or object to provide relevant services to the user [1]. Types of context information include location, user identity, activity and time, but the body itself can be a powerful source of context information. By monitoring users' actions and postures, these systems can answer questions such as: "Is the user facing a public display?", "Is the user running?", and "What is the user doing?"

A subset of context-aware systems are *human activity recognition* systems. These typically receive input from sensors, pre-process the data, segment the data stream, extract features from the segment, feed them into a classifier and label each segment according to a set of activities the classifier was previously trained upon [39]. The data used to train a classifier can be of many types (e.g. users' routines [26], ambient sensors [263], etc.), but on-body sensors can be especially informative for this purpose, as they directly capture movements and postures.

Another category of context-aware systems is *affective computing* systems—those that recognise users' emotional and psychological states. Several modalities can be used or combined to recognise affective states, such as facial expressions, prosodic features of

speech, and electro-dermal activity, but little attention has been given to body movements and postures. The few works that use body data to recognise affect, do so in a fashion similar to affect recognition, by training classifiers on a data-driven feature set.

The two categories of context-aware systems described above present a common challenge. Both tend to take data-driven approaches, often ignoring the underlying meaning of the data. This is understandable, considering the high complexity and large number of degrees of freedom of human movements and postures. By reducing this complexity into a manageable feature set composed of aggregate metrics across multiple degrees of freedom, researchers can train classifiers that work with high precision and recall for a wide range of applications.

In this thesis, we address three questions. First, how can we capture and encode body movement data? We are interested in discovering novel methods of specifying activities and recording movements for posterior analysis. Second, how can we support the inference of deeper meanings from body movement data? We are interested of what else can be inferred by working on a higher abstraction level beyond aggregate measures of sensor data, such as the quality of movement and nonverbal cues. Third, how can we help users in communicating movement information and providing feedback to improve different movements?

1.2 Lower-Body Interaction

In terms of explicit interaction, most of the work in HCI has focused on the upper body. From early mouse and keyboard interaction, through multitouch surfaces and mid-air gestures, all the way to unconventional input, such as head and gaze gestures, researchers have proposed a multitude of ways of using our upper body for computing input. We believe that the lower limbs can also provide an effective means of interacting with computers.

This does not mean that interfaces operated by the legs and feet do not exist. In fact, they are as old as HCI itself [79]. Pedals, foot mice, balance boards, dance mats, smart trainers—research prototypes and commercial products of input devices that capture input from the lower limbs have existed in a wide range of shapes and sizes. The motivations for these works vary substantially: accessible input for users with limb impairment, playful fitness exercises, multi-dimensional simultaneous input, navigation in virtual environments, and early detection of abnormal gait patterns are just a few examples of the plethora of applications for foot input. However, because of this variety, the literature on the topic is scattered in journals and conference proceedings of Ergonomics, Accessibility, Wearable Computing, among others, with no single resource providing an overview of the field or design guidelines for this kind of interaction.

Further, several of these works tend to be small-scale evaluations and developed from diverse backgrounds. This leads to a lack of deep understanding of the fundamentals of the human factors of foot-based interaction, such as pointing performance models, the effects of direction of movement, and the effect of foot dominance.

The second direction in which we investigate the research space of body movements is towards the lower body, and to do so, we set three goals. Our first goal is to rigorously consolidate the history and state-of-the-art of foot-based interaction. Our second goal is to fill the gaps in the understanding of human performance with the feet. Our third and final goal is to expand the application domains that can benefit from support from the feet as an additional input modality.

1.3 Multimodal Interaction

The plethora of input devices now in the market has made it possible to combine interaction modalities in a way that was not possible before. Such systems can treat different modalities as *redundant*, offering flexibility in the input choice (e.g. certain tablets allow users to input text with an on-screen touch-sensitive keyboard or through a peripheral keyboard attachment) or as *complementary*, combining their inputs to create a richer user experience (a process called multimodal fusion).

The additional degrees of freedom provided by multimodal input offer a solution for complex applications that require the simultaneous manipulation of multiple parameters. Examples of such applications are performances of electronic music, graphic design and 3D user interfaces. In this thesis, we believe that incorporating unconventional body-based input modalities, such as gaze and feet, into multidimensional tasks can create novel user experiences. This, however, raises many questions: How can the eyes support the hands in complex tasks? In which ways such applications can benefit from foot control? How can we bring all of these modalities together?

The third direction we explore in this thesis is towards multimodal interaction. Our first goal is to understand how the eyes can support the hands in complex tasks. The second goal is to develop novel ways in which the feet can support the hands. Finally, our third goal is to explore the kinds of novel multimodal interactions that arise from being able to track the whole body.

1.4 Methodology

In general, our methodology is based on the design, development and evaluation of algorithms, systems and interaction techniques. However, we adapt our methodology to each particular proposition investigated in this thesis.

For our first proposition, we look at implicit interaction, specifically at how to capture, analyse and provide feedback on movement data. We focus on two kinds of implicit information: the quality of movement and nonverbal cues. Because little work has been conducted in both areas, we take an exploratory approach. We conducted several data collection studies to inform system implementations and get qualitative opinions from users. We investigate novel ways of encoding activities and providing feedback on the quality of movement by developing systems to support weight lifting activities; and investigate how to extract a higher level of abstraction of movement data in the form of nonverbal cues for Affective Computing applications.

For our second proposition, we look at lower body interaction, specifically at the fundamentals of using unconstrained foot movements to support desktop interaction. Because of lack of formal knowledge on the topic, we take a focused empirical approach, driven by experiments on the human factors of foot interaction. Our first goal was to address the problem of the dispersed literature. To do so, we conducted an extensive survey of foot-based interactions, under the lenses of users, systems and interactions. We used this survey to summarise performance studies of foot-based interaction, to compile a dictionary of foot gestures, and to understand how to capture input from the feet. To address the gap in the understanding of the human factors of foot-based interactions, we built a Kinect-based foot tracker and conducted a series of experiments that quantify several aspects of this modality.

For our third proposition, we look at multimodal bodily interaction, specifically at leveraging the richness of human body movements for complex tasks. To create novel interactions that use the whole body, we take an application-driven approach, focused on developing techniques for specific tasks in the domain of 3D user interfaces. We investigate how the eyes can support mid-air gestures in 3D selection and propose foot-based

interaction techniques for the canonical 3DUI tasks. Finally, we bring both upper and lower body interaction together in a full body sensing arcade machine.

1.5 Contributions

The work presented in this thesis makes the following contributions:

- **A formalisation of the concept of quality and qualitative recognition**. We explore how quality of movement can be extracted from users' performances of weight lifting activities and how that can be communicated back to the user.
- **Systems and algorithms for capturing, encoding, analysis and feedback of body movements and postures.** We built and evaluated data-driven and model-driven solutions for assisting weight lifters and for annotating movement and postures for Affective Computing.
- **The first extensive literature review on foot-based interactions.** We surveyed related works that take input from the feet and analysed them under three lenses: human factors, sensing mechanisms and types of interactions.
- **A characterisation of unconstrained foot movement for input in a desktop setting.** We achieve this through four empirical user studies in which we derive Fitts's Law models amongst other quantitative metrics and qualitative insights about how users use free foot movement to interact with computers. To enable these studies, we built a Kinect-based foot tracker.
- **Novel upper- and lower-body interaction techniques for 3D user interfaces.** We contribute a selection technique that combines gaze and hand gestures and an exploration of how the four canonical 3DUI tasks can be performed by the feet.
- **Explorations on full body interactions for bodily play**. We built a multimodal sensing platform in the form of an augmented arcade machine (Arcade+) and developed games that demonstrate the novel interactions and game mechanics that arise from being able to capture input from the whole body.

1.6 Thesis Roadmap

This thesis is structured in five chapters. Chapters 2 and 3 focus on extracting implicit information from body movements. Chapters 4 and 5 investigate the lower body for explicit interaction. Chapter 6 explores different use cases for full body interaction in the domain of 3D user interfaces. Figure 1 shows a mind map that illustrates the topics investigated in this thesis.

- **Chapter 2: Modelling Body Movements for Activity Analysis and Feedback** explores systems to support Weight Lifting activities. In this domain, the quality of movement execution is crucial for the exercises to have their desired effects, but in activity recognition, quality is a concept that is often overlooked. We define and discuss quality in activity recognition, and explore how we can support the communication of movement quality. We propose three approaches for extracting qualitative information from movements. The first is inspired by traditional activity recognition systems: we train a classifier by demonstrating the right execution of the exercise and the possible mistakes. The second approach is an object-oriented framework that enables users to formally specify movement models and generate a feedback interface for the exercises. The third approach combines the advantages of these two by allowing users to demonstrate the correct execution and extracting the model automatically.

- **Chapter 3: Capturing Nonverbal Cues from Body Movements** examines the domain of emotion recognition from affective body expressions. In Affective Computing, the focus has been on modalities such as facial expressions, speech and physiological signals. The few works that attempt to recognise emotions from body movements do so in a data-driven fashion, by building classifiers that deal directly with sensor data. We propose to support the recognition process by working on a higher level of abstraction. We describe AutoBAP, a system that breaks down body movements and postures into individual nonverbal cues, which can be used to study affective body expressions from a higher level of abstraction.
- **Chapter 4: A Survey of Lower Body Interactive Systems** surveys works that take input from the legs and feet. As the literature on lower body interaction is scattered in different fields, we consolidate them into an extensive literature review. We structure our discussion on users' characteristics and how they affect the interaction; systems that capture input from the feet and provide feedback through them; and different kinds of foot-based interaction types.
- **Chapter 5: Empirical Investigations of Foot-Based Interaction** presents a series of focused studies that explore human factors and interaction techniques using free foot movement as explicit input for desktop computers. The survey in Chapter 4 uncovered a need for a fundamental understanding of foot-based interaction, which we address in this chapter. To enable our studies, we built a novel foot tracker based on a Kinect sensor mounted under the desk. We built performance models based on Fitts's law, tested the effects of the movement direction and the dominance of the foot on the performance, and investigated the simultaneous use of the two feet and the feet with the hands.
- **Chapter 6: Multimodal Body Movement for 3D Interaction** describes studies that benefit from the high expressivity and number of degrees of freedom of body movements to perform 3D interaction tasks. The studies focus on explicit interaction, one with the upper-body and one with the lower-body. In the first one, we combine gaze and mid-air hand gestures to enable natural 3D selection and in the second, we explore different ways in which the feet can perform the canonical 3D interaction tasks: navigation, selection, manipulation, and system control. Finally, we bring both upper and lower body sensing together in the form of an augmented multimodal arcade machine.

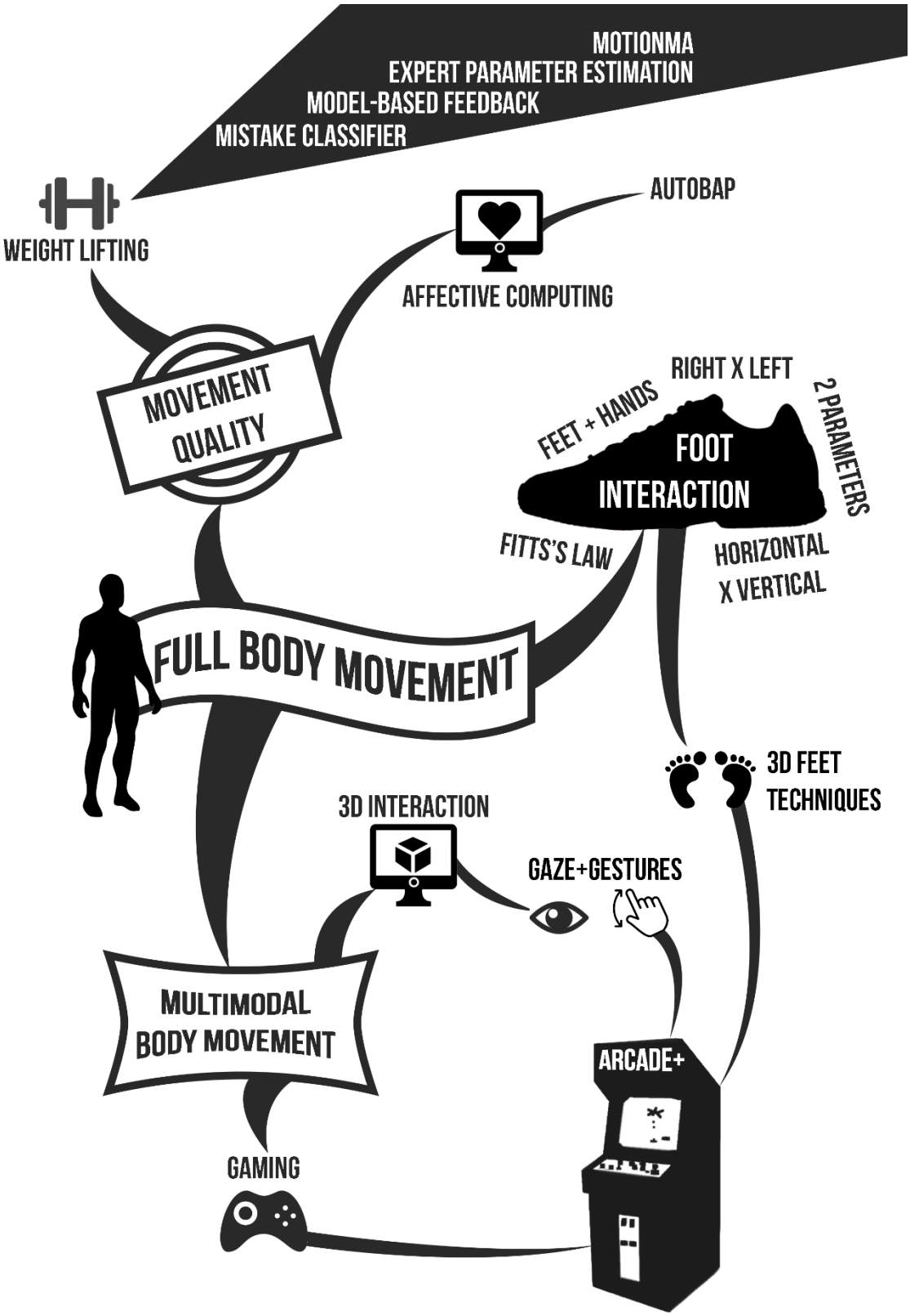


Figure 1 - This thesis mind map illustrated the themes and studies investigated in this thesis.

2 Modelling Body Movement for Activity Analysis and Feedback

"Your body is the church where Nature asks to be reverenced."

Marquis de Sade

In many areas, executing movements in a very specific way is crucial, ranging from the millimetric finger movements of surgeons to the highly expansive full body performances of gymnasts. In such cases, computers can offer a powerful way of analysing the quality of movements and providing feedback to users. Whereas substantial work has been conducted in observing movements to infer *which* activity is being performed, little work has looked at inferring *how* it is being performed. Therefore, the first proposition that we investigate in this thesis is that there is more to be inferred by users' natural movements and postures, and in this chapter we focus on their quality. However, in the context of activity recognition, the concept of activity quality is not entirely clear. If we are to analyse the quality of movements, we must first define what we mean by quality and understand what its consequences for activity recognition are. We propose a definition for the quality of an activity as the *adherence of the execution of an activity to its specification*. This definition highlights the fact that in order to measure quality, some benchmark must be specified. However, given the high complexity of body movements, this presents a considerable challenge. Moreover, the quality of a movement can often be improved with continuous feedback from experts, so computing systems should strive to support this communication process. In this chapter, we propose several methods of modelling movements for quality analysis and of supporting the communication of movement information for feedback provision.

To investigate this problem space, this chapter focuses on the domain of Weight Lifting activities. We chose this domain for several reasons. First, there has been an increased interest in augmenting sports activities, in the form of mobile phone applications and fitness bands that monitor users' steps and activities. Second, people from all ranges of experience perform these movements, so novices can strongly benefit from qualitative feedback on

their execution. Third, these exercises are usually performed with a determined set of equipment and in a controlled environment, offering great potential for augmenting the experience with sensors.

If computers are to analyse movements, the first challenge lies in capturing this information. How to best sense movements: with wearable sensors attached to users' clothing and equipment, or with remote sensors that observe performances unobtrusively? Once this information has been captured, how to specify the movement for quality recognition: should users explicitly pre-specify the parameters of the exercise or should systems learn from observing users' performances? Finally, how to provide feedback information to the user?

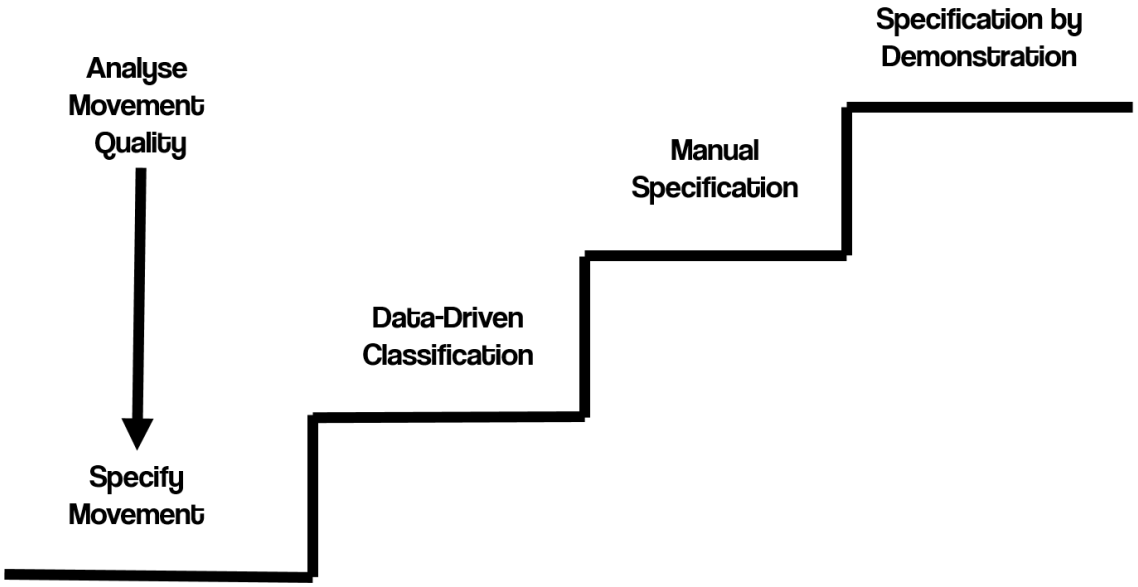


Figure 2 - Chapter outline. We set out with a goal of analysing movement quality. Our definition of quality highlights a need for movement specification and we explore three ways of accomplishing this: (1) data-driven classification; (2) manual specification; and (3) specification by demonstration.

To address these questions, this chapter begins by overviewing related work (Section 2.1) and describing the domain of the qualitative assessment of weight lifting activities (Section 2.2). We then present several approaches to modelling and analysing movement quality. Inspired by work on activity recognition, we first describe a data-driven approach in which users train a classifier by showing correct and incorrect ways of performing exercises (Section 2.3). In conventional activity recognition, algorithms look at a sensor stream and output a label that corresponds to *which* activity is being performed. We use a similar technique, but our classes reflect different *ways* in which the same activity can be performed, rather than different activities. In a user study, we demonstrate that we can successfully classify executions with high precision and recall.

However, due to the unpredictable number of possible mistakes that users can make, this approach does not scale up well. Therefore, for our second approach, we built an object-oriented, model-based framework for manually specifying exercises (Section 2.4). Our framework allows users to translate instructions from weight lifting books into activity models that can monitor different body parts and provide feedback on the movement quality by generating a feedback interface for it. We conducted two user studies to evaluate this approach. In the first study, we demonstrate that given a good model, the system can successfully assist users in improving their performance. In our second study, however, we show that experts find it difficult to explicitly estimate the parameters of the movements, such as the speed and the range of motion.

To address this problem, we finally present a system that combines the advantages of both approaches, by still using an underlying model, but learning the model parameters by demonstration (Section 2.5). We built a system prototype called MotionMA, which allows experts to demonstrate the correct execution of the movement, adjust the model parameters manually if needed and automatically create a feedback interface that monitors the performance. We show that our system can model a wide variety of movements and assist users in correcting mistakes.

2.1 Related Work

2.1.1 Recognition of Sports Activities

A large number of researchers have investigated means to provide computational support for many sports activities. For example, Michahelles et al. investigated skiing and used an accelerometer to measure motion, force-sensing resistors to measure forces on the skier's feet and a gyroscope to measure rotation [174]. Ermes et al. aimed to recognize several sports activities based on accelerometer and GPS data [80]. In the weight lifting domain, Chang et al. used sensors in the athlete's gloves and waist to classify different exercises and count training repetitions [46]. More recently, the Microsoft Kinect sensor has been used in research and uses a depth camera to extract a skeleton [72], which shows great potential for tracking sports activities unobtrusively.

2.1.2 Qualitative Assessment

Whereas several works explored how to recognize activities, only few addressed the problem of analysing their quality [39]. Haven et al. used cameras for tracking spine and shoulders contours, in order to improve the safety and effectiveness of exercises for elder people [106]. Moeller et al. used the sensors in a smartphone to monitor the quality of exercises performed on a balance board and provided appropriate feedback according to its analysis [177]. Similarly, *Wii Fit* is a video game by Nintendo that uses a special balance board that measures the user's weight and centre of balance to analyse yoga, strength, aerobics and balance exercises, providing feedback on the screen. With the objective of assessing the quality of activities Hammerla et al. used Principal Component Analysis to assess the efficiency of motion, but focused more on the algorithms rather than on the feedback [103]. Strohrmann et al. used inertial measurement units installed on the users' foot and shin to analyse their running technique, but didn't provide feedback to the user [257].

2.1.3 Model-Based Activity Recognition

Because sports exercises are often composed of well-defined movements, it is worth analysing approaches that leverage the capabilities of a model to analyse activities. For example, Zinnen et al. compare sensor-oriented approaches to model-based approaches in activity recognition [297]. They proposed to extract a skeleton from accelerometer data and demonstrated that a model-based approach can increase the robustness of recognition results. In a related work, Zinnen et al. proposed a model-based approach using high-level primitives derived from a 3D human model [298]. They broke the continuous data stream into short segments of interest in order to discover more distinctive features for Activity Recognition. Reiss et al. used a biomechanical model to estimate upper-body pose and recognize every day and fitness activities [220]. Finally, Beetz et al. used a model-based system to analyse football matches in which players were tracked by a receiver that triangulated microwave senders on their shin guards and on ball [17].

2.1.4 User Feedback

Works that offer feedback to the athlete include displaying performance statistics on a screen for rowing, table tennis, and biathlon training [8]. Iskandar et al. proposed a framework for designing feedback systems for athletes [125]. Hey et al. used an enhanced table tennis practice table to visualize past impact locations by tracking the ball using a video camera and a vibration detector [108]. A few works explored how to provide feedback on swimming technique using a GUI [181] and a multimodal approach [9]. Several works aimed to track exercises to provide feedback and thus increase motivation. Examples include the commercial *Nike + iPod* that combines data gathered from sensors in the user's shoes with music, *MPTrain* that builds a playlist by using the mapping between musical features, the user's current exercise level and the physiological response [190], and *MOPET* that uses GPS, acceleration and heart rate data to increase motivation and provide advice to the user through a 3D avatar on a mobile device [40]. There has also been work on using sensors to provide physical activity energy expenditure, since the amount of calories burnt in an exercise is a very important metric for performance evaluation. Approaches in this direction include *SensVest*, a wearable device to record physiological data from children playing sports [143] and using artificial neural networks to estimate energy expenditure [189].

2.1.5 Remote Coaching

Remote collaboration has been extensively explored in the field of Computer Supported Cooperative Work (CSCW). Previous work related to physical activities includes virtual reality systems that put the user and trainer side by side for tai chi learning and dancing [151]. There's also been work on how to convey gestures remotely using voice combined with a projection [141] and a head-mounted camera and a near field display installed on users' helmets [121]. Video recordings of trainers performing exercises have been used extensively in several different mediums, ranging from video tapes to online streaming. More recently instructors have been able to recommend sets of exercises remotely using a wide range of smart phone apps and services like *Fitocracy* and *Fitlink* that enable online collaboration among users and between users and trainers. These systems, however, can only go as far as routine prescription, without any means for users to assess their performances.

2.1.6 Motion Tracking and Analysis

In the sport sciences, a common method for analysing performance of exercises is to film the athletes and annotate the footage offline using a video digitisation system. An alternative is to use a motion tracking system to extract a skeleton of the athlete automatically. Such systems can be vision based, usually with passive markers (e.g. Vicon, OptiTrack) or IMU (Inertial Measurement Unit)-based (e.g. XSens). More recently, depth cameras such as the Microsoft Kinect and the ASUS Xtion enabled consumer-level motion tracking applications, including fitness games that guide players and give feedback on their performance, such as *Nike+ Kinect Training*, *Your Shape: Fitness Evolved* and *EA Sports Active 2*. A drawback of these commercial systems is that their algorithms are hidden from the user and their exercises are pre-programmed, without the possibility to tailor the exercises to the user's needs. Technogym's strength machines can be augmented with the *IsoControl* hardware that provides feedback on range of motion, speed and resting time as well as repetition counting. No similar product exists for free weight lifting (i.e. using dumbbells and barbells). Several research projects focused on using inertial measurement units to track and automatically assess physical exercises. For example, Chang et al. were able to tell weight

lifting exercises apart using accelerometers on users' bodies but did not analyse the quality of individual executions [46]. Moeller et al. used the sensors in a mobile phone to analyse and provide feedback on exercises performed on a balance board [177]. The Kinect has also been used to improve the quality of movements. *Kinerehab* analyses users' performances of physical rehabilitation movements and provides feedback, but the exercises only included lifting both arms to the front, to the side and upwards [47]. Martin et al. developed a system that sets out to perform a real-time ergonomic analysis of industrial workers carrying and lifting objects in order to prevent musculoskeletal disorders [167]. Their system was limited to analysing static positions rather than dynamic movements or gestures. Moreover, neither system allows users to specify the movements by demonstration.

2.1.7 Programming by Demonstration

As the feedback loop between trainers and experts is based on the demonstration of the movements, we sought inspiration in the field of *Programming by Demonstration (PbD)*. PbD has been an active area of research since the early 80's [98]. Instead of hard cording a system's behaviour, PbD aims to make it possible to program systems by having a user demonstrate to them how they should behave. Such systems aim at making their programming easier for the end-user, who does not need to learn a formal language to specify the system's behaviour [57,156]. Application areas include robotics [41], software for children [58], text editing, gesture recognition [162], children's toys [217], and context-aware computing [68]. In this work, we use a PbD approach to model human movement and assess its quality of execution. Researchers in robotics extensively studied training of robots with human demonstrating certain activities [130,244]. Hence, while previous works focused on using PbD for reproduction, prediction and recognition of movement, we use it to specify a model and analyse further performances of the same movement.

2.2 Understanding the Problem Domain

In this section we discuss the concepts of quality in the domain of activity recognition, as well as approaches to communicate movement and quality information in the domain of weight lifting. Our goal is to provide a basis for our approaches in a way that reflects weight lifters' practices, but grounded on a solid understanding of what quality is and how it can inform system design.

2.2.1 Defining Qualitative Activity Recognition

In order to discuss qualitative activity recognition we first must define what we mean by the "quality of an activity". Although a few works in activity recognition explored aspects of quality there is still no common understanding in the community as to what defines the quality of an activity and particularly what is "high" or "low" quality.

The term "quality" has been widely discussed in other fields, such as management research. The International Standards Association defines quality as the "*degree to which a set of inherent characteristics fulfils requirements*" [250] and Crosby defines it as "*conformance to specifications*" [55]. What these definitions have in common is the fact that one starts with a product specification and a quality inspector measures the adherence of the final product to this specification. These definitions make it clear that in order to measure quality, a benchmark is needed to measure the quality of a product against, in this case its product specification. Adapting this idea to the qualitative activity recognition domain it becomes clear that if we can specify how an activity has to be performed we can measure its quality by comparing its execution against this specification. From this, we define quality as *the adherence of the execution of an activity to its specification*. As a corollary, we define a qualitative activity recognition system as *a system that observes the user's execution of an*

activity and compares it to a specification. Hence, even if there is not a single universally accepted way of performing an activity, if a manner of execution is specified, we can measure its quality.

Based on the definition of quality and qualitative activity recognition it is worth discussing which are its main aspects and challenges. Qualitative activity recognition differs from conventional activity recognition in a distinctive way. While the latter is concerned with recognising *which* activity is performed, the former is concerned with assessing *how* (well) it is performed. Once an activity is specified, the system is able to detect mistakes and provide feedback to the user on how to correct these mistakes.

This directly raises three important questions. First, is it possible to detect mistakes in the execution of the activity? Traditional activity recognition has extensively explored how to classify different activities. Will these methods work as well for qualitative assessment of activities? The second question is how we specify activities. Two approaches are commonly used in activity recognition: a *sensor-oriented* approach, in which a classification algorithm is trained on the execution of activities and a *model-oriented* approach, in which activities are represented by a human skeleton model. The third is how to provide feedback in real-time to improve the quality of execution. Depending on how fast the system can make the assessment, the feedback will either be provided in real-time or as soon as the activity is completed. Real-time feedback has the advantage of allowing the user to correct his movements on the go, while an online system might make use of more complex algorithms and provide useful information without distracting the user.

2.2.2 The Importance of Physical Activity

It is well-agreed among physicians that physical activity leads to a better and longer life. For example, a recent consensus statement from the British Association of Sport and Exercise Sciences showed that physical activity can reduce the risk of coronary heart disease, obesity, type 2 diabetes and other chronic diseases [188]. Moreover, a recent study estimated that at least 16% of deaths could be prevented by improving people's cardio-respiratory fitness [25]. An effective way of improving cardio-respiratory fitness is to regularly perform muscle strengthening exercises. Such exercises are recommended even for healthy adults as they were shown to lower blood pressure, improve glucose metabolism, and reduce cardiovascular disease risk [188].

A key requirement for effective training to have a positive impact on cardio-respiratory fitness is a proper technique. Incorrect technique has been identified as the main cause of training injuries [87]. Moreover, free weights exercises account for most of the weight training-related injuries (90.4%) in the U.S. [138]. The same study states that people using free weights are also more susceptible to fractures and dislocations than people using machines. The predominant approach to prevent from injuries and provide athletes with feedback on their technique is personal coaching by a professional trainer. While highly effective, the presence of a trainer may not always be possible due to cost and availability. Personal supervision also does not scale well with the number of athletes, particularly among non-professionals. This highlights the great potential for automated analysis of such movements.

2.2.3 The Communication Process in Weight Lifting Training

Physical exercises, such as in sports or physiotherapy, require a specific execution to result in the desired training effect. Hence, experts such as personal trainers and physiotherapists need to communicate this knowledge to novices so that they can perform these movements properly. From informal observations and interviews with trainers we found that the

communication between experts and novices can usually be described by a 3-step communication loop (see Figure 3): The expert first demonstrates how the movement should be performed and gives hints on what the novice should focus on.

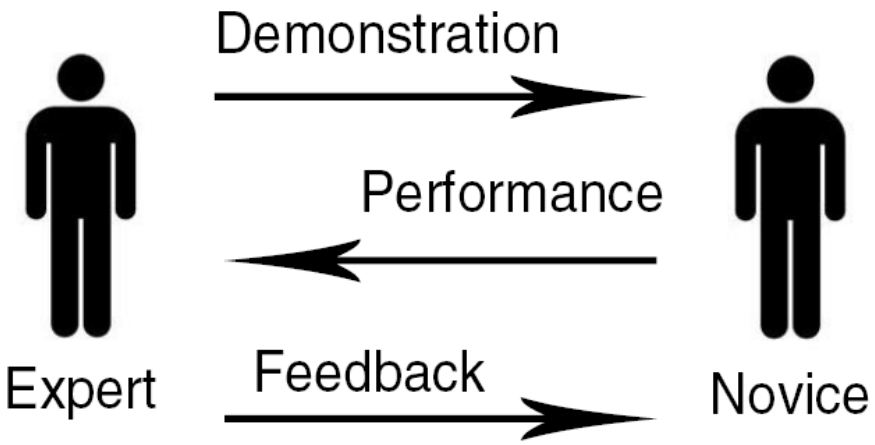


Figure 3 - Bidirectional communication loop between expert and novice. The expert demonstrates a movement, which is repeated by the novice and improved according to the expert's feedback.

This communication loop works well if the novice is under direct supervision by the trainer, i.e. if both of them are co-located, but it breaks if personal supervision it not possible, e.g. at home. In these cases, novices have to rely on video recordings or on written descriptions and images of the exercises. Whereas training using such descriptions is possible, this approach does not allow for real-time feedback and prevents novices to learn how close their execution is to the desired one and how they can improve it. In addition, written descriptions are typically high-level and qualitative and do not allow novices to quantitatively analyse their performance. In this chapter, we look for ways of supporting this communication process through an automated system.

2.3 Detecting Mistakes using a Data-Driven Approach

The goal of our first experiment was to assess whether we could detect mistakes in weight-lifting exercises by using data-driven activity recognition techniques. We recorded users performing the same activity correctly and with a set of common mistakes with wearable sensors and used machine learning algorithms to classify each mistake. From the point of view of our definition of activity quality, we used the training data as the activity specification and the classification algorithm as the means to compare the execution to the specification.

2.3.1 Participants and Apparatus

We recruited six male participants aged between 20-28 years, with little weight lifting experience. For data recording we used four 9 degrees of freedom Razor inertial measurement units (IMU), which provide three-axes acceleration, gyroscope and magnetometer data at a joint sampling rate of 45 Hz. Each IMU also featured a Bluetooth module to stream the recorded data to a notebook running the *Context Recognition Network Toolbox* [11]. We mounted the sensors in the users' glove, armband, lumbar belt and dumbbell (see Figure 4). We designed the tracking system to be as unobtrusive as possible, as these are all equipment commonly used by weight lifters. We made sure that all participants could easily simulate the mistakes in a safe and controlled manner by using a relatively light dumbbell (1.25kg).

STUDY AT A GLANCE

Goal: Record data to train a mistake classifier

Method: Data recording

Participants: 6M (20-28y.)

Procedure: 10 Biceps Curl reps in 5 different ways

Results: 98% recognition performance with a 2.5s window

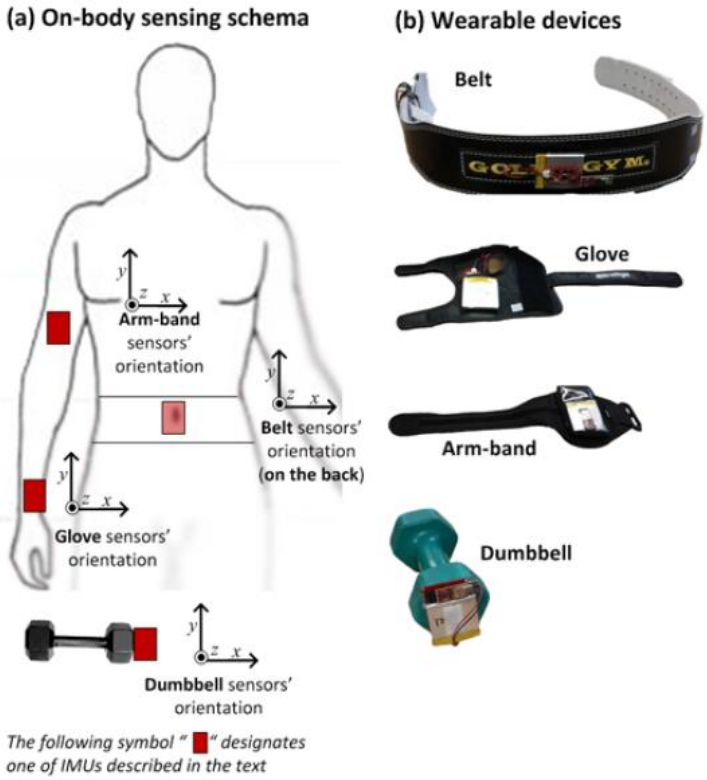


Figure 4 - Sensing setup, with inertial measurement units attached to conventional weight lifting equipment.

2.3.2 Procedure

Participants were asked to perform one set of 10 repetitions of the Unilateral Dumbbell Biceps Curl in five different fashions: exactly according to the specification (Class A), throwing the elbows to the front (Class B), lifting the dumbbell only halfway (Class C),

lowering the dumbbell only halfway (Class D) and throwing the hips to the front (Class E). Class A corresponds to the specified execution of the exercise, while the other 4 classes correspond to common mistakes. Participants were supervised by an experienced weight lifter to make sure the execution complied with the manner they were supposed to simulate.

2.3.3 Feature Extraction and Selection

For feature extraction we used a sliding window approach with different lengths from 0.5 second to 2.5 seconds, with 0.5 second overlap. In each step of the sliding window approach we calculated features on the Euler angles (roll, pitch and yaw), as well as the raw accelerometer, gyroscope and magnetometer readings. For the Euler angles of each of the four sensors we calculated eight features: mean, variance, standard deviation, max, min, amplitude, kurtosis and skewness, generating in total 96 derived feature sets.

In order to identify the most relevant features we used the feature selection algorithm based on correlation proposed by Hall [101]. The algorithm was configured to use a “Best First" strategy based on backtracking. 17 features were selected: in the belt, were selected the mean and variance of the roll, maximum, range and variance of the accelerometer vector, variance of the gyro and variance of the magnetometer. In the arm, the variance of the accelerometer vector and the maximum and minimum of the magnetometer were selected. In the dumbbell, the selected features were the maximum of the acceleration, variance of the gyro and maximum and minimum of the magnetometer, while in the glove, the sum of the pitch and the maximum and minimum of the gyro were selected.

2.3.4 Recognition Performance

Because of the characteristic noise in the sensor data, we used a Random Forest approach [33]. This algorithm is characterized by a subset of features, selected in a random and independent manner with the same distribution for each of the trees in the forest. To improve recognition performance we used an ensemble of classifiers using the “Bagging" method [32]. We used 10 random forests and each forest was implemented with 10 trees. The classifier was tested with 10-fold cross-validation and different windows sizes, all of them with 0.5s overlapping, apart from the 0.5s window. The best window size found for this classification task was of 2.5s and the overall recognition performance was of 98.03% (see Table 1). The table shows false positive rate (FPR), precision, recall, as well as area under the curve (AUC) averaged for each of the 5 tested on 10-fold cross-validation over all 6 participants (5 classes). With the 2.5s window size, the detailed accuracy by class was of: (A) 97.6%, (B) 97.3%, (C) 98.2%, (D) 98.1%, (E) 99.1%, (98.2% weighted average).

Table 1 - Recognition performance

Window Size	FPR	Recall	AUC	Precision
0.5s	3.9	85.0	97.4	84.9
1.0s	1.8	93.5	99.5	93.5
1.5s	1.0	96.5	99.8	96.5
2.0s	0.7	97.2	99.9	97.2
2.5s	0.5	98.2	99.9	98.2

We also conducted a leave-one-subject-out test in order to measure whether our classifier trained for some subjects is still useful for a new subject. The overall recognition performance in this test was 78.2 %. The result can be attributed to the small size of the datasets (approximately 1800 instances each dataset, extracted from 39.200 readings on the IMUs), the number of subjects (6 young men), and the difficulty in differentiating variations of the same exercise, which is a challenge in Qualitative Activity Recognition. The use of this approach requires a lot of data from several subjects, in order to reach a result

that can be generalized for a new user without the need of training the classifier. The confusion matrix of the leave-one-subject-out test is illustrated on Figure 5.

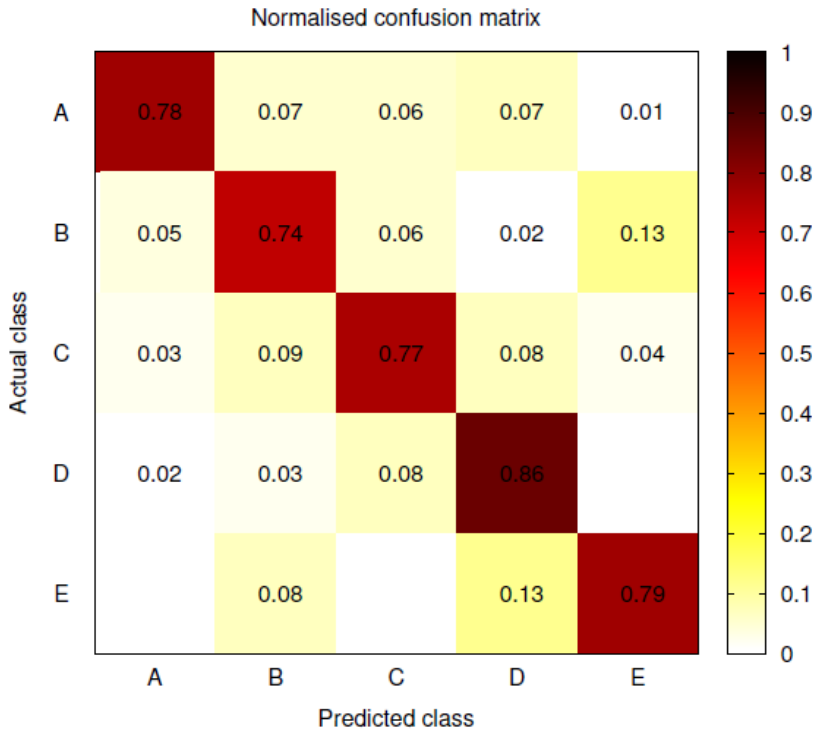


Figure 5 - Summed confusion matrix averaged over all participants and normalised across ground truth rows.

2.3.5 Discussion

The advantage of this approach is that no formal specification is necessary, but even though our results point out that it is possible to detect mistakes by classification, this approach is hardly scalable. It would be infeasible to record all possible mistakes for each exercise. Moreover, even if this was possible, the more classes that need to be considered the harder the classification problem becomes.

2.4 Specifying Exercises with a Model-Based Approach

Due to the inherent problems of the classification approach, we concentrated our efforts into formalising a way of specifying activities and recognizing mistakes by looking at deviations from the model in the execution. This section outlines our second approach to qualitative activity recognition systems for weight lifting that helps minimize the effort of translating specifications into systems. We implemented a C# framework for the development of such applications tracking body movements with a Kinect sensor. The following sections illustrate the conceptual steps of our approach on the example of building a feedback system for the Unilateral Dumbbell Biceps Curl and the Unilateral Lateral Dumbbell Raise exercises using our framework. We evaluate whether they can assist users in correcting mistakes in a user study. Finally, we evaluate how well expert weight lifters can estimate the parameters of the movement to configure the specification.

2.4.1 Activity Selection

An activity must have an appropriate granularity to be analysed. If the activity is too complex, it is more appropriate to break it down into smaller activities. In our example, even though a weight lifting exercise is commonly performed in sets of 6-12 repetitions, for our purposes we consider an activity as a repetition of the exercise. This way we can analyse each repetition separately. A Biceps Curl repetition involves raising and lowering the dumbbell, so we define the beginning of the activity as when the user starts to lift it and the end as when it reaches the initial position again.

2.4.2 Activity Specification

The activity should be specified as clearly as possible in natural language. The clearer the specification is the easier it will be to model the activity. In our example, we used as the specification the instructions provided by a weight lifting book [285]. An activity specification can be comprised of several instructions. For the Unilateral Dumbbell Biceps Curl, the specification we used, adapted from Williams et al. [285], was the following: (1) Stand solidly upright; (2) Your feet should be shoulder-width apart; (3) Your shoulders should be down; (4) Curl the dumbbell in an upward arc. Curl the dumbbell to the top of the movement when your biceps is fully contracted; (5) Elbows pointing directly down and return to the start position; (6) Don't lean back and throw your hips to the front.

2.4.3 Activity Modelling

For each instruction in the activity specification we create a model of the recognition mechanism according to the components in the framework. The components can be of five different classes: Joint, Operator, Feature, Counter and Classifier. The model architecture is illustrated in Figure 6. Figure 7 shows an example of an instruction modelled accordingly.

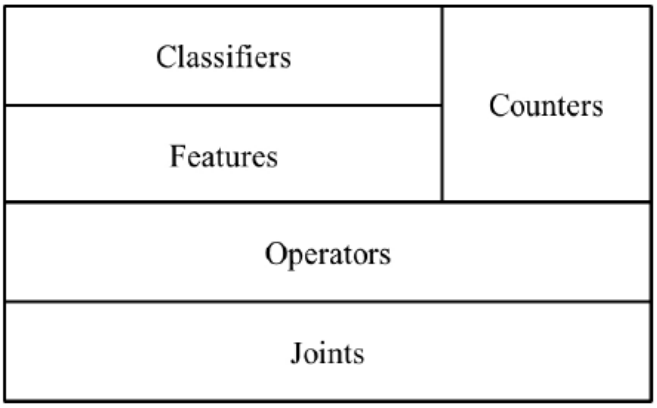


Figure 6 - Layered architecture of our model, which receives as input the raw position of the joints as provided by a tracking system and outputs a class of quality for the instruction

A Joint in our model is the XYZ position of each of the 20 joints provided by the Microsoft Kinect 1.0 Beta2 SDK. Different instructions will make use of different sets of **Joints**. In the example in Figure 7, the joints are the Wrist, the Elbow and the Shoulder of each side.

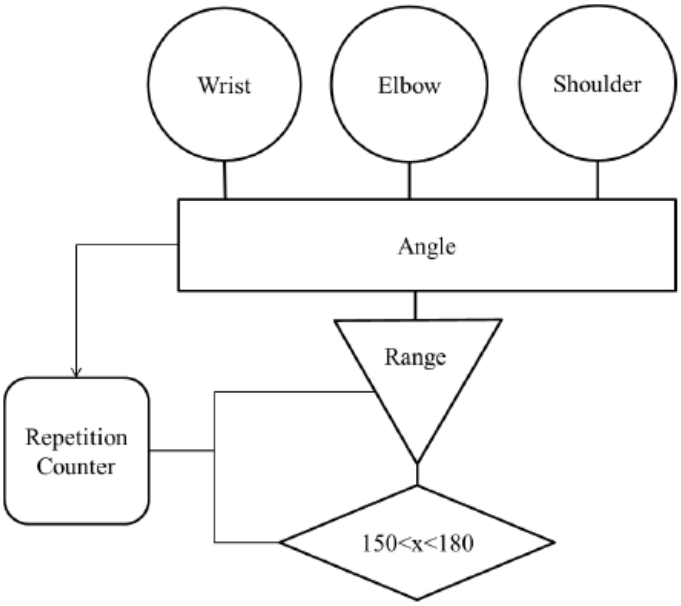


Figure 7 - Example of an instruction specification based on our model. From the joints' position coordinates, we extract the angle between them and its range to count repetitions. For each repetition, we calculate the overall range and check whether it is within the specified limits.

Operators represent operations performed on top of the raw position coordinates of a single joint or a set of joints. The implemented operators include the XYZ coordinates, distance and angles between joints. For example, in the modelling of Instruction 4, we could describe the movement in terms of the trajectory of the hand, but this wouldn't be ideal because it would depend on the length of each user's arms. Hence, we use an Operator to convert it to the angle between the Hand, Elbow and Shoulder instead, because this is not a user-dependent measure. **Feature** components buffer the data that is provided by the operators and perform statistical analyses (such as mean, standard variation, range, energy, etc.) on a dataset when an event is triggered. In the example, because we want to make sure the movement is complete, we measure the range of the angle.

The classification is triggered by **Counters**. In our approach, we can classify an exercise in two ways: continuously (with features being sampled in short intervals) or discretely (with features being sampled after every repetition). If you need a feature to be monitored after a specified time interval, you can use the **Clock Counter**. If you want the feature to be extracted for each repetition, you can use a **Repetition Counter**, which triggers events after detecting a repetition. Finally, the classification of the quality of the execution of the instruction is performed by **Classifier** components. These can range from performing simple thresholding operations to running more complex machine learning algorithms. In Figure 4, the Angle between the Wrist, Elbow and Shoulder is fed into the Repetition Counter, which uses a strong filter and a peak counting algorithm to detect repetitions. When a new repetition is detected, this component trigger the calculation of the range.

Once the model is complete, the class library we implemented allows the programmer to translate directly the components in the model into an object-oriented application. All that is required is to input the parameters in the instantiation of the components and to connect the components by subscribing to each other's events. We modelled and implemented the feedback systems for 3 exercises: Unilateral Dumbbell Biceps Curl, Unilateral Dumbbell Triceps Extension and Unilateral Dumbbell Lateral Raise.

We tried to keep the Classifiers as simple as possible so they could be easily tweaked on the spot. This is useful for a real life scenario where the trainer might want make minor alterations in the specification. For example, a general specification for the Biceps Curl says that one should curl the Dumbbell all the way to the top. However, it is possible that the trainer might want the athlete to perform the exercise only halfway to the top in order to stimulate specific muscle fibres. Our system is prepared to allow these parameter modifications to be made easily.

2.4.4 Parameter Adjustment

There will be times when the available instruction is more qualitative than quantitative, so some instructions should be adjusted and parameterised to account for that. For example, one of the instructions for the Biceps Curl was to “Curl the dumbbell in an upwards arc towards your shoulder". This instruction does not provide the metrics to unambiguously build the model. One possible interpretation is: the angle between the wrist, the elbow and the shoulder should go from 180 to 0 degrees. However, these values need to be tested and adjusted to make sure they correspond to the measurements provided by the Kinect SDK. The framework allows you to debug this step using events that let you monitor each step of the analysis.

2.4.5 User Feedback

In the user interface, the system should give feedback for the conformance to each one of the instructions in the specification separately. The feedback should be as clear as possible using different visual and auditory cues. The classifiers output different classes of quality that can be translated into traffic lights that would turn green if the specification was OK and red in case of problems in the exercise, for example. Because of the complex nature of the exercises, it is also recommended to give feedback on how to improve the technique. The focus of this specific work was not on comparing different feedback approaches, but, rather to develop a complete pipeline of analysis that was able to analyse movements. Nevertheless, we implemented a feedback interface as shown in Figure 8.

2.4.6 Evaluating the System

We carried out a user study to evaluate a system developed using our framework to check whether our approach to qualitative activity recognition can lead to improvement in the quality of exercise performance.

2.4.6.1 Procedure

STUDY AT A GLANCE

Goal: Test the effect of feedback on weight lifting performance

Method: Within-subjects experiment

Participants: 8M (20-28y.)

Procedure: 3 x 10 reps of 2 exercises with 3 different weights

Independent variable: presence of the feedback interface

Dependent variable: mistake count

Results: 23.48% fewer mistakes in the Lateral Raise and 79.22% in the Biceps Curl

First, participants were asked to fill in a questionnaire regarding their experience with weight lifting prior to the execution of the exercises. The 8 participants were all male, 20-28 years old, with little or no experience in weight lifting. The feedback systems include a traffic light that indicates whether an instruction is being performed correctly and messages instructing the user on how to improve the execution. They also featured a range of motion indicator and a repetition counter. The user could see himself performing the exercise, with the feed from the camera built in the Kinect sensor. The interface is illustrated in Figure 8. Then the participants were asked to perform the Unilateral Biceps Curl and the Unilateral Lateral Raise. We wanted to compare the execution of these exercises with and without the feedback system, so we provided them with a written description of the expected execution of the exercise and asked them to perform each exercise with a hand while the feedback system was turned off. Then, we turned the feedback system on and asked them to perform the same exercises but now with another hand, in order to minimize the effects of tiredness. Each exercise was performed in three sets of ten repetitions with increasing weights of 1.25kg, 3.0kg and 7.0 kg. Participants were instructed to stop whenever they felt uncomfortable. We recorded data using a Kinect sensor connected to a Windows 7 PC. The feedback was provided using a 27-inch LCD display. After the exercises, participants were asked to fill in another questionnaire regarding the experience with the feedback system.

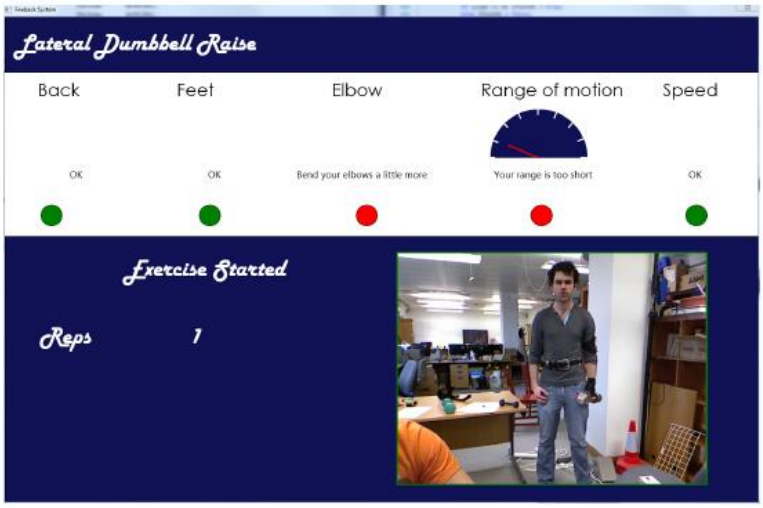


Figure 8 - User interface and feedback system for the Unilateral Dumbbell Lateral Raise exercise

2.4.6.2 Results

With the Lateral Raise feedback system users made 23.48% fewer mistakes per repetition, while with the Biceps Curl feedback system users made 79.22% fewer mistakes. Participants were rank ordered by the number of errors in each of the two conditions. A Wilcoxon matched pairs signed ranks test indicated that the number of errors was significantly lower when using feedback (median = 4.5) than without using feedback (median = 26.5), $Z = 2.52$, $p = .008$, $r = .63$. For the lateral curl exercise a Wilcoxon matched pairs signed ranks test indicated no significant difference between the two conditions, $Z = 1.61$, $p = .125$. These results indicate great potential for such systems in correcting mistakes and consequently improving the quality of weight lifting activities.

Table 2 - Questionnaire results

Question	Mean	Standard Deviation
How helpful do you think such system is in a gym environment?	4.57	0.53
How clear was the presentation of information?	4.14	0.90
How much do you believe the feedback influenced your performance?	3.86	0.90
Did you try to change your movements according to the feedback?	4.71	0.49
Did the feedback improve your performance?	3.57	0.79

Table 2 shows the mean and standard deviation of the questionnaire results averaged over all participants. Values correspond to responses on a 5-point scale with 5 representing strongly positive and 1 strongly negative answers. User responses were generally positive regarding doing the exercises with the aid of a feedback system. Some suggestions for improvement include making the messages larger and easier to read and trying out different feedback visualizations, like video or 3D animation.

2.4.6.3 Discussion

Our results showed significant improvement in the Biceps Curl. The Biceps Curl is fairly well known exercise, which people perform without taking the time to think about the technique, so the system worked well in aiding users correcting mistakes. Only a small improvement was detected for the Lateral Raise. Even though users made almost a quarter of mistakes made without the system, we can't say that the results are statistically significant. This points out to a potential in the system, but further inspection is necessary. We attribute this result to the difficulty of performing this exercise with the provided weights. The Lateral Raise stimulates mainly the deltoids, which are significantly weaker muscles than the biceps, so a fall in performance was expected. Users were generally very positive about the system. Some claimed to be "more conscious about movements due to both the camera image and feedback visualisation" and to feel "more confident in the movements I was making and able to correct mistakes." The use of feedback systems was praised: "Without the feedback system you cannot be sure whether you are doing the exercise properly", indicating that this field of research deserves more attention. Some participants thought the simple interface was good ("simple signals gave exact instructions on what to correct") while others had some suggestions on how to improve it ("red and green could be avoided (...) as colour blind people will not be able to see the difference" and "I would prefer images that illustrate what to improve"), showing that how to design interfaces for such systems is a challenge on its own.

2.4.7 Evaluating Experts' Ability to Estimate Exercise Parameters

Our model-based approach has the advantage of formalising exercise instructions in a normative way—you specify the correct execution and the model searches for deviations of this specification. On one hand, this frees the user from having to demonstrate mistakes as in our first approach, but on the other hand, it requires users to specify the quantitative parameters of the movement (e.g. the range of motion and the duration of each repetition). To investigate how well experts are able to formalise a movement description from their knowledge and experience, (i.e. to specify and estimate angles and speed of movements), we conducted a user study.

STUDY AT A GLANCE

Goal: Test experts' abilities in estimating exercises' parameters

Method: Data recording

Participants: 10M (22-45y.)

Procedure: 10 reps of 3 exercises with 3 different weights. We also asked for their estimates for the parameters in their performance

Results: Estimates are, on average, 11˚ off from the actual parameters.

For this study, we chose three common weight lifting exercises, all of which novices were able to perform and that experts were able to instruct others on how to perform. The exercises were the Unilateral Dumbbell Biceps Curl, the Unilateral Dumbbell Lateral Raise and the Unilateral Dumbbell Triceps Overhead Extension. We recruited 10 male expert weight lifters, aged between 22 and 45 years (mean = 31.0, std = 8.38), heights ranging from 1.70 m to 1.89 m and weights ranging from 63.0 kg to 90.8 kg. Participants were recruited using posters distributed around the university campus and on two gym clubs. We made sure that all participants had at least three years of experience with weight lifting, ranging from that up to 25 years.

The experiment took place in a quiet indoor laboratory setting. Participants were asked to wear a belt, armband, gloves and hold a dumbbell on each of which were mounted 5 Xsens inertial measurement units as shown on Figure 9. These sensors were connected by wires in a daisy chain fashion to an Xsens XBus, which streamed the data over Bluetooth to a Linux laptop. Data recording and synchronisation was handled by the Context Recognition Network Toolbox (CRNT) [11]. Participants were asked to stand approximately 2 meters in front of a Kinect camera. A custom application streamed the skeleton data to the CRNT, as well as the labels and the frame number of the RGB camera. As ground truth we recorded videos using a camcorder mounted to the right side of the participants.

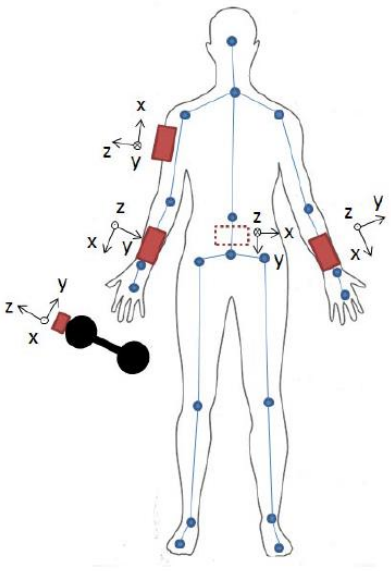


Figure 9 - Sensing setup. Blue circles indicate the joints tracked by the Kinect and red rectangles indicate the position and orientation of the Xsens sensors.

Upon arrival in the lab all participants were asked to sign a consent form and to answer a brief questionnaire regarding their previous experience with the specific exercises and weight lifting in general. They were then presented with a written description of the three exercises. Afterwards, they were guided through a structured interview by the experimenter, in which they were asked to provide the angles for certain joints of the body for each step of each exercise. Specifically, our goal was to obtain the values of angles that they considered to be ideal for each movement in their own understanding of how these movements should be performed when exercising or coaching as defined in the written description. We were interested in finding for each joint, whether there was movement during the exercise, the initial and final angle, the acceptable tolerance and how long the movement should take. These are all standard measures in Kinesiology [22]. Since we are focusing on upper body movements, we inquired about arm flexion/extension/hyperextension, arm abduction/adduction, arm lateral/medial rotation, arm horizontal abduction/adduction, elbow flexion/extension, wrist pronation/supination, wrist flexion/extension and ulnar/radial deviation. To make it clear to the interviewee what each movement meant, we provided a graphical diagram of each movement along with the corresponding angular axes. Also, participants were free to ask questions if in doubt about what each movement meant.

After equipping our sensing system participants were asked to perform 10 repetitions of the Unilateral Dumbbell Biceps Curl, the Unilateral Dumbbell Lateral Raise, and the Unilateral Dumbbell Triceps Extension using three different weights (1.25kg, 3kg and 7kg), totalling 90 repetitions for each participant. The data was manually annotated according to the exercise, weight and participant for post-hoc analysis.

2.4.7.1 Results

We analysed the data by manually separating each repetition according to the plots and the frames in the video recordings. From each repetition, we extracted different information depending on the exercise. For the Biceps Curl, we extracted the maximum and minimum of the elbow flexion, the mean of the arm flexion, the mean of the arm abduction, the mean of the lateral rotation and the duration of each repetition. For the Lateral Raise, we extracted the maximum and minimum of the arm abduction, the mean of the arm flexion, the mean of the lateral rotation, the mean of the horizontal abduction, the mean of the elbow flexion and

the duration of the repetition. Finally, for the Triceps Extension, we extracted the mean of the arm flexion, the mean of the arm abduction, the mean of the lateral rotation, the mean of the horizontal abduction, the minimum and maximum of the elbow flexion and the duration of the repetition.

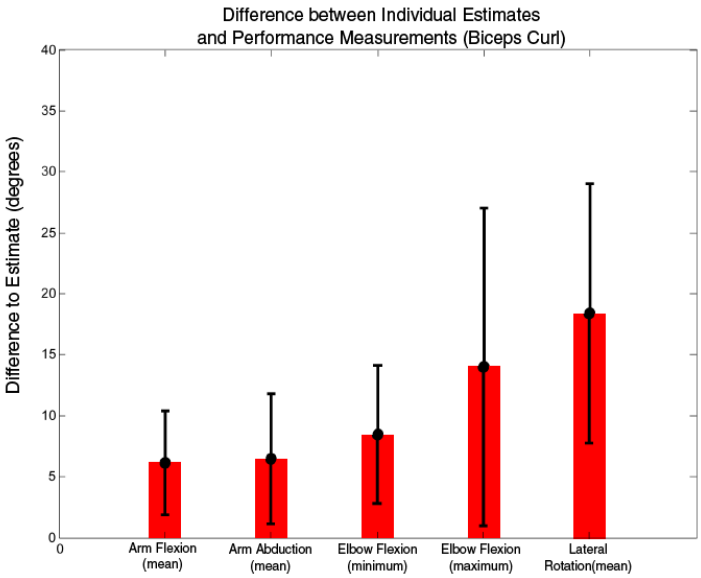


Figure 10 - Average difference among all participants of each angle against the value suggested by the corresponding participant for that angle. The error bars represent standard deviations. For angles that are supposed to remain still, we use the mean and for those that are supposed to vary we use the maximum and minimum.

For each participant, we compared the distribution of measurements for each of these values to the answer given in the questionnaire with a one-sample t-test. With very few exceptions, the angles suggested in the interview were significantly different ($p < .05$) to the angles in the actual performance. This means that participants' estimates of the angles don't reflect their performance. Figure 10 shows the consolidated differences for all participants between each of the measured angles in the 30 performances of the Biceps Curl and the estimate given by the corresponding participant. Differences of the same order of magnitude were found for the other exercises. Figure 11 shows the average difference for all angles in all movements for each participant.

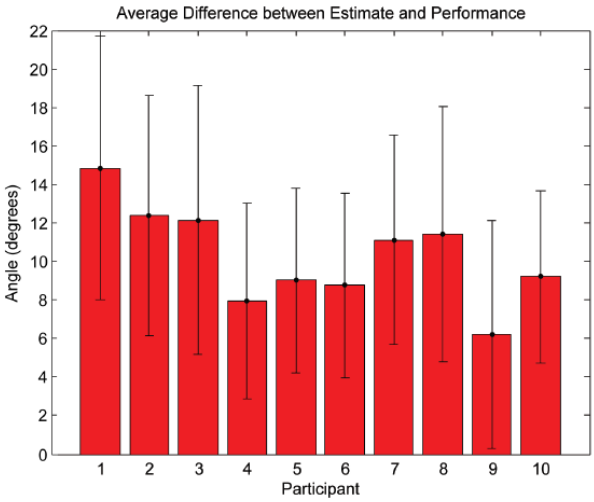


Figure 11 - Average difference for each participant between the estimate and the measured angles among all angles. This study shows that users' estimates are on average 10.76 degrees off the measured angle, making it hard to rely on these estimates to evaluate performances.

In order to find out whether the performances matched among different participants, we ran a One-Way ANOVA Test on each dataset and again the results were significantly different ($p < .05$), even though in a some cases the performance of a few different trainers would match within the group, as seen on a post-hoc Tukey analysis.

2.4.7.2 Discussion

The results suggest that even experienced weight lifters find it difficult to give an accurate estimate of the angles in their own performance, even for movements that are simple and very well defined. Indeed, it has been long known that humans tend to overestimate acute angles and underestimate obtuse ones [187]. This is evidence that specifying a movement by natural language and estimating precise angles by observation is difficult.

Moreover, even though they all knew the movements rather well from previous experience and were provided with a written description of their execution, performances varied significantly among different participants. This does not mean, however, that some of them did the exercises incorrectly. It highlights that a wide variety of small variations in the performance of the same movement are possible, which is evidence for the ambiguities of written descriptions of movements. Also, these variations occur according to the goals of the weight lifter. For example, whereas the standard description of a Biceps Curl might instruct to start the movement with your arm fully extended and to finish the movement with your arm fully flexed, a trainer might ask you to flex your arm only halfway to exercise specific muscle fibres. This does not make the second execution incorrect, only different to the first one, which indicates the need to be able to encode these variations into descriptions in a clear way.

2.5 Mediating Communication with a Modelling by Demonstration Approach

If even experienced users have difficulty in explicitly specifying movement angles, we can assume that results would be even worse when considering a wider range of experiences and movement complexities. Moreover, even if they accurately provide this information, inputting these values into the system would prove to be a tiresome and demanding task,

which reinforces the case for a more implicit way to obtain this information. Inspired by the observation of how sports trainers and physical rehabilitation professionals communicate movement in real-world scenarios, we developed an approach to extract the movement model by having the user demonstrate it. We called our system *MotionMA*.

2.5.1 MotionMA

MotionMA is a system to communicate movement information in a way that people can convey a certain movement to someone else who is then able to monitor his own performance and receive feedback in an automated way (see Figure 12). It allows users to model movements and assesses further performances of the same movement in comparison to the model.

In developing a system to support the communication of movement information, we tried to emulate the process described on Figure 3. This implies three main functionalities that the system should make available: (1) Allow users to specify movements; (2) Analyse performances of movements; (3) Provide feedback on performances.

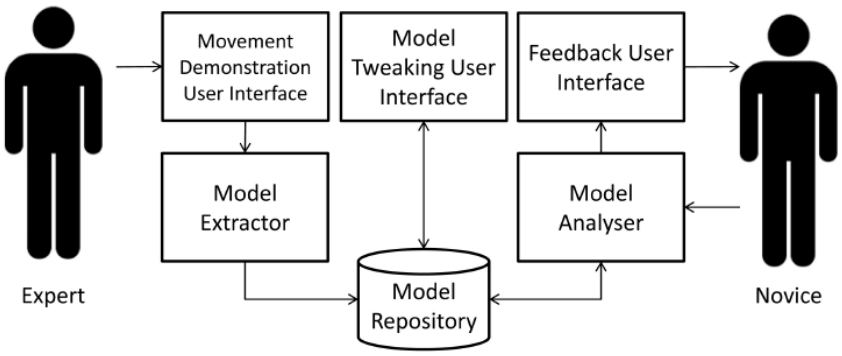


Figure 12 - MotionMA system architecture. Experts specify movements through the demonstration interface. The system extracts a model of the movement and stores it in a repository. The models can be visualised and edited in the tweaking interface. Novice users can then perform the movement, which is then compared with the model to provide feedback accordingly.

All three functionalities are tightly coupled to each other. The feedback provided by the system depends on the output of the analysis, which, in turn, strongly depends on the movement's model. At the core of the system is the movement's internal representation, i.e. its model. We wanted our model to be simple enough to be suitable for real-time analysis, but be able to convey enough detail to provide an accurate description of the movement. We also wanted it to be expressive enough to model a wide variety of movements and to be subject-independent, so that the same model can be used by different users. Moreover, we wanted it to allow for analysis approaches that provide relevant information, such as mistake spotting and improvement guidelines.

The standard in Kinesiology is to define motion in terms of the angles of each bones in relation to reference planes (sagittal, frontal and horizontal), which makes it suitable for people with different body measurements. In practice, however, we don't need all three angles, since the direction of a vector can be specified in 3D space with only two. Therefore, for each bone we define a spherical coordinate system with the origin at one of the joints and the zenith direction as the vertical at the global coordinate system. Our model is defined as a set of timestamped characteristic points for each bone in each rotation axis. A set of

characteristic points is a minimum collection of points with which you could generate an approximation of the time series. The equation below shows the formal representation of our model, where the models $M_\theta$and $M_\varphi$for each one of the 19 bones b in a skeleton S are the sets of n tuples(t_{θ_i}, θ_i) and m tuples$(t_{\varphi_i}, \varphi_i)$ with the timestamped values of the polar θ and azimuthal φ angles.

$$\forall b \in S \begin{cases} M_\theta(b) = \{(t_{\theta_1}, \theta_1), \dots, (t_{\theta_n}, \theta_n)\} \\ M_\varphi(b) = \{(t_{\varphi_1}, \varphi_1), \dots, (t_{\varphi_n}, \varphi_n)\} \end{cases}$$

We don't use tuples with both angles at once, so n can be different than m. This allows us to analyse the movement in each axis independently, which makes it easier to provide improvement guidelines. For example, if a problem is detected at the azimuthal axis, we can guide the user to rotate the corresponding bone up or down, whereas if the problem is at the polar axis, the rotation would be in the left or right direction.

Using this model, we can compare two performances by comparing corresponding sets of points. First we need to compute the distances between each point in one set to each point in the other to match the points. By looking at each component of the difference vector, we can then infer whether each characteristic point was at the correct angle (y axis) at the correct time (x axis). Moreover, depending on the direction of this vector and the rotation axis, we can infer improvement guidelines. For example, if the difference between a point in a given performance and the corresponding point in a baseline performance for the azimuthal angle is positive in the x axis and negative in the y axis, we can infer that the user got that point too soon and at an angle too large. The magnitude of each component will tell how far off was he from the specification and whether an improvement guideline should be displayed. Also, we can tell genuine mistakes and normal variability by looking at the distribution of values in the demonstration and checking whether the value in the performance differs significantly from the ones in the demonstration.

2.5.2 Modelling Movement by Demonstration

Our approach to specifying the movement model draws inspiration from Programming by Demonstration (PbD) works. Our system allows the movement to be specified by allowing the user to demonstrate it, analogous to how professionals do in reality. While this minimises users' effort in inputting information, it creates new challenges, such as detecting the beginning and end of the activity, detecting the characteristic points of the movement and accounting for variations in the movement.

A problem shared by many PbD and Activity Recognition systems is segmenting the actual activity being demonstrated, i.e. detecting when it begins and when it ends. In order to solve this problem, we take as input for our model extraction algorithm a dataset with the performance of 10 repetitions of the movement. Once we can detect each repetition, we can trim the demonstration data at either end.

Detecting the repetitions, however, is not a trivial task. Chang et al. counted repetitions of free-weights exercises by applying a strong low-pass filter and extracted features to train a Hidden Markov Model for individual exercises [47]. Because at demonstration time there is no previous data to be used for training, a machine learning approach is unsuitable for our goal. Instead, we ask the user to perform 10 repetitions of the movement and assume that we will find 10 cycles of a pattern in the data. Because during the movement some bones might be static while others move, we pick one axis of a bone to count the repetitions in all of them. We choose the bone and axis by counting peaks and valleys in every axis of every bone and looking for a dataset that gives us 10 repetitions with the largest amplitude of angles. The peak counting algorithm uses a strong low pass filter combined with an autocorrelation algorithm that looks for zero-derivatives on the data, with the mean of values as a threshold to eliminate noise and small variations. Once we have the repetition

separation, we automatically have the duration of the movement, which we use in the analysis algorithm to analyse speed.

In order to find the characteristic points of the movement, our algorithm analyses each repetition separately. We do this by using a weaker low pass filter in every dataset and by looking at zero-derivatives in the data. This gives us peaks, valleys and inflexion points, which works well to provide a general shape of the curve. At the end of this step, we have 10 sets of points that describe the same movement. We merge these sets to get a consolidated model. We accomplish this by looking for the centroids in the merged data using a k-means clustering algorithm. By using a consolidated model, we account for variations in individual repetitions of the movement. With this set of points at hand, we can make the distinction between static and dynamic axes and tag them accordingly. This information will be used in the analysis to analyse each dataset appropriately. Figure 13 shows the complete process for an example dataset.

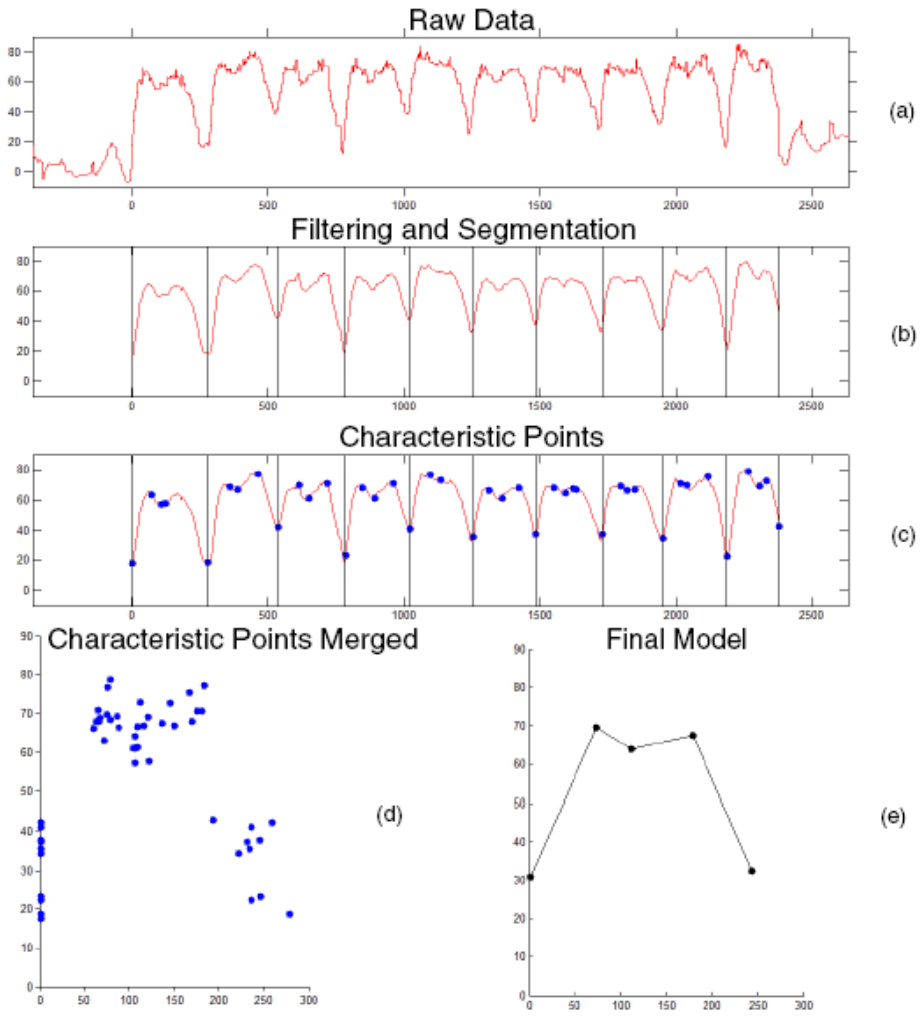


Figure 13 - Model extraction from demonstration performance. The raw data (a) is filtered and by counting repetitions, the data is segmented (b). We then find the characteristic points for each repetition (c), merge them (d) and look for the centroids of the data clusters (e).

The demonstration user interface is shown in Figure 14. The main application was programmed in C# and receives the data from the Kinect sensor directly through its SDK

(version 1.5). From the main menu, the user selects 'Demonstrate' and is presented with the demonstration interface. Here he can see his own image in a virtual mirror overlaid with the skeleton tracked by the Kinect sensor. The user then stands in the starting position and issues a voice command to start the recording. He proceeds to perform 10 repetitions of the movement as consistently as possible. When he is finished he issues another voice command to stop the recording. The system then uses the extraction algorithm described in the previous section to generate a motion model. The user can then save or discard the generated model. The demonstration modelling is done in *Matlab*, using a COM automation interface to transfer data between the two applications.

Figure 14 - Demonstration interface. The user can see his skeleton overlaid on top of the colour image recorded by the Kinect. The recording is controlled by voice commands.

In order to correct eventual mistakes in the extracted model and to improve its overall accuracy, the user can choose to tweak the model. The 'Tweak' interface is shown in Figure 15. The system displays a skeleton figure on which the user can select the bone he wishes to visualise. When a bone is selected, the system displays the plots of the model for each axis of the bone as well as whether each axis is dynamic or static. In this step, the user also selects which bones he would like to see in the feedback interface. This allows users to tailor the system to specific goals, such as balance or range of motion. He can then save or discard the changes.

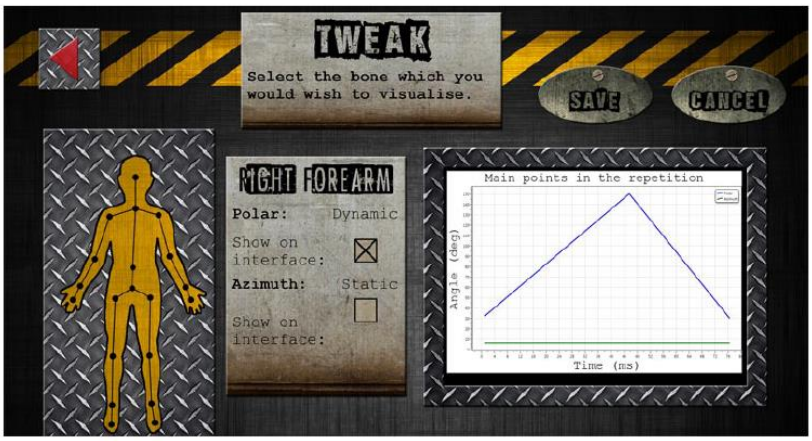


Figure 15 - Tweaking interface. In this step, the user can visualise the extracted model for each bone and select what is to be monitored in the performance interface.

In the example of the biceps curl, a user could use this step to make sure the model for both arms are the same as well as adjusting the range of motion in the y-axis. Also, he could make sure that the duration of the flexion and extension during the movement are correct by adjusting the position of the characteristic points in the x-axis.

2.5.3 Model Analysis and Feedback

Once a motion model is extracted from a demonstration, we can use it to analyse further performances of this movement. Since the relevance of the feedback is tightly coupled to the analysis outputs, we analyse each performance in different levels to allow for different kinds of feedback. Even though further research is necessary into how to actually provide this feedback, we tried to make our analysis approach as comprehensive as possible.

The input for our analysis algorithm is the same as in the demonstration, i.e. the spherical coordinates of each bone in the skeleton as well as the model extracted at the demonstration step. We analyse the data at three separate points in time, giving feedback accordingly: continuously, at the end of a repetition and at the end of a set of repetitions. Continuous feedback is given for rotation axes that are supposed to remain static. This means that the angles at these axes should remain still around a predetermined value, so they are monitored continuously looking for deviations that are flagged as soon as they are found. Dynamic axes are monitored at the end of each repetition. These datasets require the complete analysis, which is described in the introduction. The system buffers the values from the skeleton and runs the analysis when it detects the completion of an instance of the movement. After the user completes all repetitions, the system can analyse the dataset as a whole and provide a more detailed evaluation of the performance due to the reduced attention overload. In this work, we focus on the real-time feedback, i.e. the continuous analysis for static axes and the repetition analysis for dynamic bones.

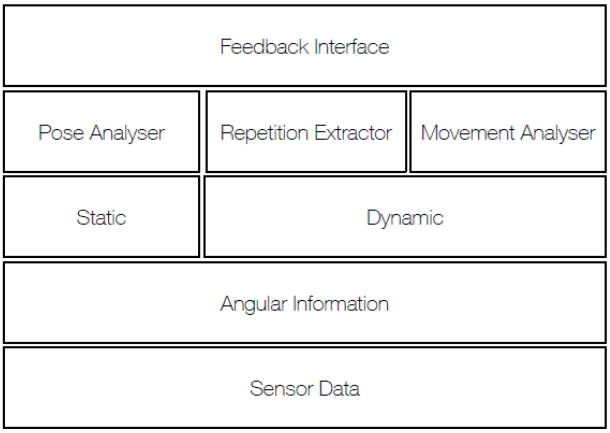


Figure 16 - Analysis architecture. Data from the tracking system is converted to spherical coordinates, analysed by the appropriate component and the final analysis is displayed on the interface.

The architecture of the analysis is comprised of the elements shown in Figure 16. The raw data from the tracking system is converted into spherical coordinates and depending on whether the dataset is static or dynamic it is analysed by a Pose Analyser or a Movement Analyser, respectively. The Movement Analyser is triggered by the Repetition Extractor. The analysis outputs are then showed in the feedback interface.

Much more work is needed to investigate what and when to display to the user. In this first attempt to answer this question we tried out different approaches to communicate movement: video recording of the demonstration; moving dials to monitor the adequate range of motion; traffic lights to monitor static axes and to indicate how to correct them; coloured skeleton, that changes the colour of each bone from green to red depending on its score; speed warning lights, that tells the user to speed up or slow down accordingly; repetition counter, that shows how many instances of the movement have been performed. The design of these interface elements was based on our previous work, described in the previous sections in this chapter and on the feedback provided by other systems (e.g. the range of motion dial is similar to how TechnoGym's equipment displays it). The overall theme of our design was industrial-like, similar to how other products are portrayed in this market.

Figure 17 - Performance interface. Information regarding static bones is displayed on traffic lights whilst the ranges of motion of dynamic bones are displayed on dials. The user can see the video recording of the demonstration and his own skeleton as tracked by the Kinect with each bone in a different colour depending on its score. The interface also displays the repetition count and warnings when the speed is too fast or too slow.

Figure 17 shows the user interface for the feedback system we implemented. The systems loads the model, as well as the interface elements selected by the user in the 'Tweak' step. If this step was skipped, the system selects the information to be displayed automatically, based on the variation detected in the model. The user stands in the starting position and issues a voice command to start the analysis. Even though further research into how to convey movement information in a meaningful way is necessary, we took the first steps in that direction by trying out different approaches. Information regarding static axes can be visualised in traffic lights that indicate whether the pose is correct and how to correct it. Dynamic axes can be visualised in dials that move together with the performance. The system also displays warnings when the performance is too fast or too slow and displays the repetition count. Also, the system displays the video feed as recorded in the demonstration session and the skeleton of the user as currently tracked. Each bone in this skeleton is coloured differently according to its score.

2.5.4 System Evaluation

2.5.4.1 Participants and Apparatus

STUDY AT A GLANCE

Goal: Evaluate our modelling and analysis approach

Method: User testing

Participants: 10 (20-41y.)

Procedure: 10 reps of a movement of their choice. Inspection of the model and mistake detection.

Results: 70% mistake recognition, high ratings for model accuracy

With a working prototype that implements our approach at hand, we could evaluate it in a user study. The goals of this study were to evaluate the modelling and analysis. For this purpose, we recruited 10 participants aged between 24 and 41, of which one was female. They had different levels of experience with weight lifting ranging from 'none' to 'experienced'. This was to ensure that the modelling system was robust enough to handle different consistencies of performance. The study was carried out in two steps, each one of which was designed to evaluate each of our goals. This section describes our experimental procedure and results. The study took place in a quiet laboratory setting and each session was carried out with a single participant. The experimental setup consisted of a 27" display with a Kinect sensor mounted on a tripod behind it. When using the system, participants would stand on a previously marked cross on the floor, approximately 2m away from the display and sensor. The system uses the Kinect as a recording device, so no experience with gestural interfaces was needed. Even though we record participants' perceptions about the user interface, our goal was not to evaluate the effectiveness of the feedback provided.

2.5.4.2 Procedure

We first evaluated the expressiveness and accuracy of our modelling approach. Each user was asked to think of a controlled and repeatable movement which they would model using our system. Before performing the actual movement, they were asked to provide a detailed written description of it, give it a name and describe five possible mistakes or variations that a person trying to learn the movement would most likely make. They were also asked which bones could be used to count the repetitions of this movement. Then, each participant recorded their movement using the demonstration interface.

We then displayed the extracted model in the "Tweak" interface and went through the model for each axis of each bone together with the participant in a structured interview fashion. We selected each bone, and the corresponding plots of each angle would appear on the screen, as well as the values for each step of the model. The user would then fill in a qualitative questionnaire where he would rank the accuracy of the model for each axis of each bone as well as whether the system marked the bone as static or dynamic correctly. They were also asked to rank the axis and bone that the model extraction algorithm used in order to count repetitions.

When evaluating the analysis, we wanted to find out (1) whether the system could recognise a good performance and (2) whether the system could spot the mistakes foreseen by the participants. We asked them to perform 10 repetitions of the movement as closely as possible to the demonstration performance and to pay attention every element in the feedback interface. After this performance, they filled in a questionnaire regarding how accurate the system was at counting repetitions, displaying the correct range of motion in the dials, showing a green light for static bones and indicating the correct speed. Then, users were asked to do 10 repetitions of each mistake and look for elements of the interface that would spot these variations. After each performance, users were asked to fill in a questionnaire regarding how accurately the system spotted each one of them. Finally, each

participant filled in a general questionnaire regarding the overall accuracy of the model, the detection of correct and incorrect instances of the movement and the repetition counting. They were also prompted to point out positive and negative aspects of the system.

2.5.4.3 Results

Participants chose movements of a wide variety, from common strength exercises (Dead Lift, Lateral Raise, Biceps Curl) to some amusingly named body gestures (The Lawnmower, Robot Elevator, Circulation Agent). These included both upper and lower body movements and could all be considered controlled and repeatable.

When inquired about the accuracy of the model for each axis of each bone on a 5 point Likert scale, users rated it were very high (median = 5, mean = 4.7785, std = 0.7181) as shown on Figure 18.

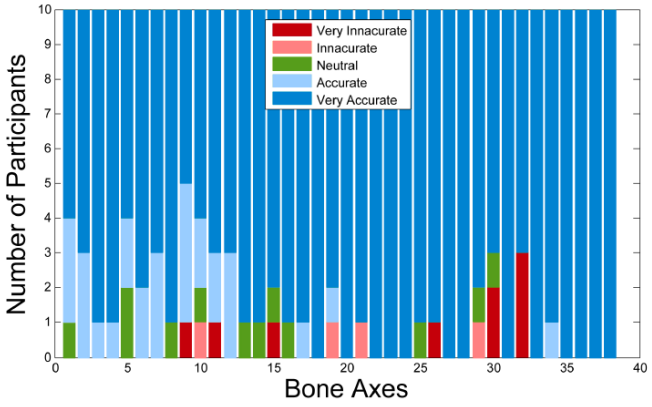


Figure 18 - Users' responses regarding the accuracy of the model. Each column represents each of the 38 axes of bone movement (2 for each of the 19 bones).

Users were also asked to rate how well the system spotted each of the 5 mistakes they came up with. Figure 19 shows users' ratings of the mistake detection accuracy for all mistakes. Figure 20 shows users' responses regarding how accurate the system was in counting repetitions, in displaying the correct range of motion, in displaying the correct posture in the traffic lights and in displaying the correct speed in the speed signs.

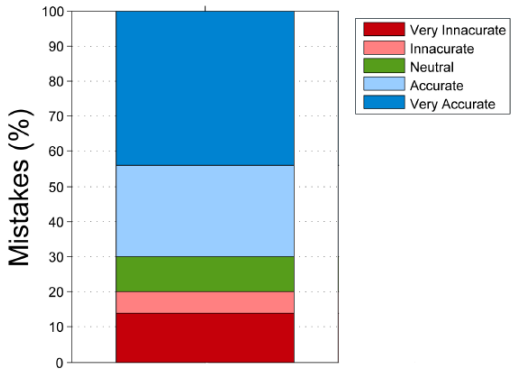


Figure 19 - Each user was asked to rate the accuracy of the mistake detection for each of the 5 mistakes they came up with. This chart show that our system accurately detected around 70% of mistakes.

In the end, users gave their overall impressions of the system. Figure 21 shows the answers regarding how much they agree that: (1) the system was able to extract an accurate model of the demonstrated movement; (2) the system was able to detect a correct performance of the movement; (3) the system was able to count repetitions accurately; (4) the system was able to detect mistakes in the movement. Most responses were positive, with the exception of the one where due to the poor choice of the repetition counting dataset, the analysis did not work as expected. The consequence of a poor choice of a repetition counting limb is that all parameters were calculated at the wrong times, making the feedback unintelligible and the repetition counting meaningless.

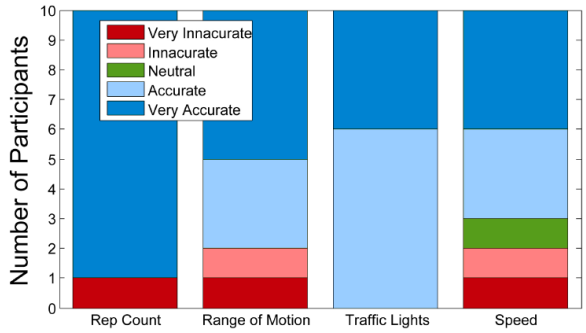


Figure 20 - Users' responses regarding how accurately each interface element would indicate a correct execution when performing the movement in the same way as in the demonstration.

Comments ranged from being very positive, recommending deployment in real world situations ("*Very impressed with the ability to analyse body and movement repetitions. I would like to see this implemented in gyms*") to negative in the case where the system did not work as expected ("*Counting repetitions was inaccurate. Feedback was confusing*."). Some participants complained about the amount of information on the screen ("*Too many things going on to look at!*"), the limitations of the tracking system ("*It cannot track hand opening/closing*.") and limitations in the algorithm ("*It didn't pick up when I was going in the wrong direction*"). Most responses, however, complimented the interface ("*I really like the interface!*"), the gamification of the movement communication ("*It looks like a cool game!*") and how the feedback could be used to correct mistakes ("*Feedback was easy to use as basis for correcting movement*").

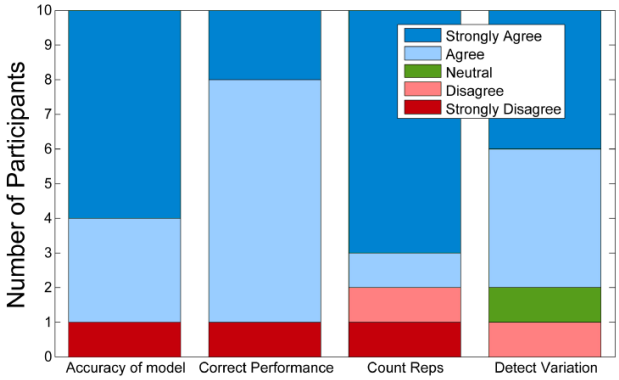


Figure 21 - Users' responses regarding their perception of the system.

2.5.4.4 Discussion

In general, participants were very positive about both the modelling and the analysis, as the high scores show. While this does not prove that the extracted model is entirely accurate, it does reflect that the extracted model makes sense in terms of ranges and general rotation of the bones to the users.

Even though users seemed to agree with the extracted model, factors such as the difficulty in interpreting plots and the previously mentioned difficulty in estimating angles make it necessary for us to obtain more evidence of its accuracy. Among the interface elements, the speed meter was the one that performed the worst. We posit that this happened due to the nature of the movement analysis. The speed is measured after a repetition of the movement is completed, by comparing how long the user took to complete it with the duration of the repetition in the model. Because the displayed speed always regards the previous repetition, we noticed that users would speed up or slow down accordingly but with no effect in the warning signs (which would only change after the repetition was completed), which generated some frustration. With the exception of one movement our repetition counting algorithm counted all 10 repetitions very accurately. The one case where it failed was due to a poor choice of the repetition counting dataset, which by consequence, produced a poor model overall.

Our results indicate that the system is able to extract an accurate model of controlled and repeatable movements and to generate automatically a feedback interface that provides feedback on the execution of the movement. The data in the user study also demonstrated some limitations in our approach. Whilst the analysis algorithm correctly identified most mistakes, there is still plenty of room for improvement. The negative cases were due to the user not being able to recognise the system's feedback, limitations in the tracking system (for example, the Kinect can't tell whether the hand is open or closed, which was an element present in some mistakes), limitations in the algorithm (for example, the analysis algorithm analyses each bone individually, so it would not detect mistakes that had to do with the relation of the movement of one bone to another).

2.5.4.5 Limitations

Although the Kinect proved to be very accurate in tracking coarse movements, when limbs were pointed directly at the camera, or occluded by the body, the overall tracking was severely penalised. The tracking of hands and feet was not precise enough to track some of the mistakes participants suggested. Also, due to the limitations of the tracking system, we limited the scope of this work to movements where the user was standing up and facing the camera, but there are a wide variety of movements in which the user is in positions that can't be tracked by the Kinect. Another limitation imposed by the tracking system is that we treat each bone as a vector in the 3D space, without taking into account the rotation of the bone around its length. This means that there is currently no support for detecting rotations such as pronation and supination of the wrist.

Another limitation regards our algorithmic approach. Our current implementation looks at the absolute orientation of the bones. Some of the mistakes participants suggested, however, in terms of the orientation of the bones in relation to one another. In future versions, we hope to add the support to model a hierarchy of bones that can be used to address this problem and suggest more valuable feedback.

Our evaluation has two main limitations. First, we only evaluated the algorithmic approach with a single user each time. Even though the system was designed for remote collaboration between multiple users, more work is required to evaluate this aspect. The second limitation is the realism of the environment. In the real world, such system would be used by athletes and trainers or patients and physiotherapists. We evaluated the system for its expressiveness by letting the users choose their own movements, but further work is

necessary to ensure that the system attends the requirements for specific application domains.

In this work, we explored the general communication process between experts and novices when transmitting movement information as described in Figure 3 and built a system that is a first step in implementing it, focusing on how experts can use it to model movements and configure feedback. Even though it was outside the scope of this particular work, there is yet a lot to be done in how effectively this information is actually conveyed to novices and how much the feedback impacts their performances.

2.6 Conclusion

We began this chapter defining quality in the context of activity recognition. Our definition highlights the importance of a specification against which to measure quality. However, given the high complexity and number of degrees of freedom in full body movement, building a specification for a movement is not trivial. For this purpose, we proposed three approaches.

Our first approach was inspired by other works in activity recognition. These works often record sensor data of users performing different activities and train a classifier that given a certain dataset, attributes a label corresponding to the activity being performed. However, instead of using different activities as class labels, our classes were the different *ways* of performing the same activity. We demonstrated that we can successfully classify different performances in this way with high precision and recall. The main advantage of this approach is that users do not need to explicitly program the system. Instead, the algorithm observes correct and incorrect executions and learn to classify further performances. This also makes it easy to give feedback to the user, as each class of mistake will have a corresponding way of fixing it. However, this approach also presents some serious issues. First, by treating the quality recognition as a classification problem, the system does not quantify quality on any continuous scale. For example, if the algorithm detects that the range of motion of a movement is too short, it does not tell the user anything about how close he is to the optimum range. Even though this could be minimised by breaking the problem further into more classes (e.g. “way too short”, “too short”, “almost there”), this makes the classification problem harder and might still not give feedback with an appropriate granularity. The second problem is that there are countless possible mistakes that users can make when executing such movements. A classification approach would require the system to be trained on each of them, and the number of classes would quickly get too large. The third problem is that by requiring the trainer to demonstrate mistakes the system effectively makes him susceptible to injuries. Whereas for a large portion of mistakes, this would not be much of a problem, certain incorrectly performed movements could easily hurt or sprain the person demonstrating the mistake. Finally, the fourth issue is that the specification is encoded into a classifier, which makes it difficult for trainers to easily make modifications. For example, if the trainer wants to recommend a shorter range of motion for a certain user, he must train the system again, even if he already has a classifier for a large range of motion.

To address these problems, our second approach relied on an explicit specification of the movement. We built an object-oriented framework that allows users to model the exercises according to common requirements for weight lifting and that could automatically create a feedback interface for it. This approach offers several advantages. First, it makes it easy to modify parameters of the movement. If the trainer requires a larger/shorter or faster/slower movement, it is as easy as changing the value of a variable. Second, it makes it easy to encode explicit knowledge from weight lifting books, as each instruction can be directly encoded into the system. Third, it allows the reuse of components, which is useful

for instructions repeated across many exercises. For example, many exercises require the feet to be still, shoulder-width apart. Once this instruction is programmed, it can be reused across all exercises. Despite these advantages, this approach also presents certain drawbacks. Our first study demonstrated that once the parameters are defined, the system leads to a better performance, however, the second study revealed that even expert weight lifters have difficulty in estimating these parameters. The system does allow for them to easily tweak the parameters, but this leads to a trial-and-error process that is hardly ideal. The final problem is that whereas the framework simplifies the encoding of *explicit* knowledge about the quality of movement, it fails in providing a platform for the encoding of *tacit* knowledge, which was nicely provided by the first approach.

To maintain the advantages of both approaches, we developed MotionMA, a system that still uses a model as in the second approach, but infers the parameters of this model by demonstration, as in the first approach. We demonstrated that our system allows users to easily create feedback interfaces for a variety of movements, even for those that are not necessarily weight lifting movements. Also, because our system uses joint angles in its model, it can be used across multiple users, supporting a communication process in which an expert can demonstrate a movement and a novice can benefit from the feedback interface generated. We envision a future in which experts all around the world can record and sell feedback interfaces based on their own performance on an online coaching marketplace that can guide users into mimicking their execution. The results from the analysis algorithms can also be used for applications that gamify the exercising experience, adding yet another layer of feedback in which users can compete to improve performance statistics in a wider variety of contexts. For example, whereas nowadays it is easy to measure and compare the distance and speed at which athletes run, no similar metrics can be automatically computed for weight lifting activities. By monitoring the movement quality, users will not only be able to compare how much weight they can lift, but also whether they can lift it with a proper technique.

3 Capturing Nonverbal Cues from Body Movement

"Every gesture is a gesture from the blood, every expression a symbolic utterance... Everything is of the blood, of the senses."

Henry Williamson

The first proposition of this thesis is that there is more to be inferred from body movements and postures. Whereas in the previous chapter, we focused on the quality of movement, this chapter looks at the emotions conveyed by our body language. Emotion is a huge component of human behaviour. Every action we take, every conversation in which we engage, every thought that springs to our minds, all are substantially shaped by our current emotional state. If emotion plays such an important part in influencing interaction between humans, it is only natural that researchers attempt to incorporate emotional aspects into Human-Computer Interaction—what is known as Affective Computing. In general, Picard defines Affective Computing as "*computing that relates to, arises from, or influences emotion*" [209]. Therefore, the focus of Affective Computing is in "*creating technologies that can monitor and appropriately respond to the affective states of the user in an attempt to bridge the communicative gap between the emotionally expressive human and the emotionally deficit computer*" [59].

Whereas modalities such as facial expressions and physiological signals have been extensively explored for this purpose, little attention has been given to affective body expressions. Only recently there has been an increased interest in using them for affect recognition, as tracking body movements now is easier than ever. Works that attempt to do so usually employ techniques analogous to activity recognition, based on training classifiers from raw sensor data.

In the previous chapter, we described several ways of working on a level of abstraction above raw data through a movement model to extract qualitative information about exercise execution. In this chapter, we use a similar approach to extract nonverbal cues for affective computing. We believe that working this way can provide significant advantages.

First, work on other modalities has successfully employed such approach to detect emotions. For example, work on facial expression analysis highly benefitted from the maturity of the Facial Action Coding System (FACS), which breaks them down into their component units. Second, working on a higher level of abstraction allows better communication of knowledge between computer science and psychology professionals, as it provides a standard for specifying movements and postures. Third, as work in Psychology finds correlations between emotions and their corresponding bodily behaviours, this knowledge can be directly injected into the recognition pipeline.

Until recently, however, there was no standard way of specifying affective body expressions in the same way of facial expressions. An attempt to overcome this problem was made by Dael et al., who proposed the Body Action and Posture (BAP) coding system [60]. However, coding BAP labels manually is a long and tiresome process. In this chapter, we use the lessons learned from our works on weight lifting analysis to automatically extract BAP annotations.

We begin with an overview of emotions and nonverbal signal annotation. We then describe the Body Action and Posture coding system and describe ways of automatically recognising emotions. Finally, we describe and evaluate our approach for extracting nonverbal cues from body data, in the form of a system prototype (AutoBAP).

3.1 What are emotions?

Despite being a crucial element in our lives, emotions have always been a difficult and fleeting concept for mankind to fully understand. In Plato's dualism, the soul is structured separately by cognition, emotion and motivation. Aristotle, on the other hand, found such separation impossible, and suggested that these components are actually interwoven [231]. Other thinkers also had different ideas about emotions. Descartes thought that a few basic emotions underlie the whole of emotional life. Darwin suggested that emotions have an evolutionary origin and are therefore universal across cultures, strongly influencing the psychobiological school of thought of emotions. Contrary to this view is that of the emotions have a sociocultural origin, a view taken by many sociologists and anthropologists. Current theories believe that both factors shape our emotional responses [231].

Different conceptualisations of emotions have been proposed in Psychology, and are usually classified as discrete or continuous models. Discrete (also known as categorical) models of emotion closely resemble how we verbally refer to them, by giving them individual names. When we talk about 'happiness' or 'sadness', we are using a categorical model of emotions. Proponents of such theories found evidence for a small number of basic emotions that are universal [267,268]. Particularly influential were the works of Ekman and Friesen who demonstrated that Fore tribesman in Papua New Guinea, a preliterate traditional society with little contact with industrial societies, both expressed and recognised the same emotional facial expressions as Western cultures [75]. These authors proposed a set of six universal basic emotions (happiness, anger, sadness, disgust, surprise, and fear), later extending it (with the addition of contempt, guilt, shame, interest, embarrassment, awe and excitement) [77]. Other discrete models include Parrott's tree structure [204] and Plutchik's wheel [210]. Criticisms of these models include the lack of correspondences between discrete emotions and brain activity, the variability in facial expressions and behaviour and impossibility of specifying gradation in emotional responses [15].

Continuous (also known as dimensional) models of emotion represent states as a tuple in a multidimensional space. In these models, instead of giving them labels, emotions are specified in a spatial continuum: anger, for example, would be represented as a high-arousal, negative-valence emotion. In Russell's Circumplex model, emotions are distributed in a two-dimensional circle, with arousal and valence in the axes [227]. Mehrabian and Russell's PAD model, specifies pleasure, arousal, and dominance as the axis [172]. Lovheim's

cube specifies emotions in terms of the release of three hormones: dopamine, noradrenaline and serotonin [161]. These models enable researchers to visualise how close or apart emotions are in space and to characterise different intensities of the same emotion. The disadvantage of these models is in that people do not normally think about emotions in terms of a multidimensional space, so it makes it more difficult to obtain ground truth from participants. However, because the emotions normally included in discrete models tend to be extreme and stereotypical, continuous models allow modelling more subtle variations in emotional states that are arguably more useful for affective systems.

3.2 Annotating Nonverbal Signals

The most widely researched modality in emotion research is facial expression. The Facial Action Coding System is a coding system that deconstructs facial expressions into action units (contractions and relaxation of facial muscles) and their temporal segments [73,74,76]. Automatic implementations of FACS include a system trained to automatically detect action units in order to differentiate fake from real expressions of pain [158] and to analyse expressions of neuropsychiatric patients [102]. Techniques to achieve this include analysing permanent and transient facial features in frontal face image sequences [266], using independent component analysis and support vector machines [51] and using Gabor wavelets with neutral face average difference [16].

An early notation system for body motion was Labanotation [152,153], which was originally developed to describe dance movements and is part of Laban Movement Analysis, which breaks movements down to Body, Effort, Shape and Space. DMAR [218] offers a graphical interface for dance experts to annotate dance concerts or clips, but it does not do it automatically. Birdwhistell's coding system is based on linguistic principles [24]. It defines kinemes (analogous to phonemes in Linguistics), which are groups of movements which are not identical, but communicate the same meaning. This notation has been used to categorise the emotions in emoticons [214]. Attempts to facilitate the transcription of body movements include animating a 3D skeleton to annotate arms' gestures [184], but this system still requires the annotator to match the animation to the video recording.

More recently, Dael et al. proposed a coding system for the description of body movement on anatomical, form and functional levels, more suitable for coding nonverbal emotion expression [60]. Some advantages of this coding system are that it minimises observer bias by being supported by a reliable observation protocol; because it is not based on linguistic principles it is independent from other modalities such as speech; and it is generic enough to be used outside of emotion research. Presently, the authors perform the coding using the Anvil software, which is a manual annotation tool [140]. As of yet, there is no system that extracts the BAP coding automatically from body motion. In Section 3.4, we give more details on BAP.

3.3 Automatic Recognition of Affective Body Expressions

Most of the research on affect recognition focuses on facial expressions, prosodic features of speech or physiological signals (see Zeng et al. for a survey of methods [293]). As motion capture systems become more affordable and widespread, increasing attention has been given to the role of body expressions in affect recognition [137,142]. De Gelder offers twelve reasons for considering body expressions for affect recognition, including the fact that bodily expressions are recognised as reliably as facial expressions [90].

Works on recognizing affect from body expressions can be distinguished depending on whether they are data-driven, or knowledge-driven; use acted, elicited, or natural

expressions; use communicative, functional, artistic, or abstract movements; and on the affective states they attempt to recognize.

Data-driven approaches use the output of a motion capture system to select relevant features and train a classifier. An example is the work of Kapur et al., who recorded data using a Vicon motion capture system and trained different emotion classifiers with techniques such as logistic regression and decision trees [135]. Knowledge-driven approaches not only draw from results in Human Sciences to assist classification, but output results that contribute to this body of knowledge. Such works usually derive some labels from the raw data using notation systems such as Labanotation [229] or the Body Action and Posture Coding System (BAP) [61].

The nature of the affective expression can be acted, elicited or natural. Acted data is usually recorded from actors and the ground truth is considered the expression they were asked to perform. The quality of the data can be enhanced by using emotion elicitation processes derived from Psychology and Theatre methods, such as Stanislavski's System [173]. Natural data is harder to record and annotate, but provide more realistic portrayals of emotions. While there has been a recent trend to-wards recording natural data, the number of such datasets is still small [137].

Affective states can be manifested in different types of movement [137]. Communicative movements are directly related to the expression of affect. Functional movements are the ones used to perform tasks and may be modulated by affect. Artistic movements are usually choreographed and exaggerated and abstract movements are used neither to communicate a meaning nor to accomplish a task.

The affective states observed in previous work differs widely. Whilst early work tended to focus on trying to cover several emotions, recent work has moved towards understanding different qualities of fewer states (e.g. Griffin et al. distinguish types of laughter from body expressions [94]) or the states conveyed by specific movements (e.g. Gunes and Pantic classify different states using only head movements [97]).

3.4 The Body Action and Posture Coding System

BAP separates its behaviour variables into 12 categories: head orientation, action and posture; trunk orientation, action and posture; arms action and posture; whole body posture; gaze; action functions and other. In BAP, orientation labels are coded using an external frame of reference, namely the interlocutor. Posture labels can be of three types: (1) posture units (PU), which are broken down into (2) posture transition phase (PT) and (3) posture configuration phase (PC). Posture transition refers to the period of time to reach the end position and posture configuration is the period of time in which the posture is maintained. The direction of postures is coded according to the three orthogonal planes that cross the centre of mass of the body in the standard anatomical position. Actions change more frequently than postures, so they are coded differently as action units, which can be broken down into its different steps (action subunits).

Even though the coding system attempts to be as objective as possible, it offers challenges to its automatic implementation. The first challenge is about how the data is segmented. The coding guidelines are very specific about the definition of onset and offset points for the segments, so it is important to define precisely the frames where the label begins and ends. However, when implementing it automatically, there will always be issues of noise and synchronisation between different sensors due to different sample rates. Second, even if the segmentation is correct, its labels depend on its context, so the same data segment may have different labels depending on the previous and next segment. Let's take the example of a right head turn. If after turning the head the user holds the head to the right, it is labelled as a posture transition, but if it is followed by a left head turn, it is labelled as an action sub-

unit. Moreover, if the user's head posture was already annotated as being turned to the right before turning the head to the right, the head turn is considered part of the configuration phase and not labelled at all.

3.5 AutoBAP

Facial expressions and speech are rich sources of information and powerful modalities for automatic recognition of basic affective states and have been investigated for a long time in affective computing research [78,82]. With the availability and decreasing cost of ambient and on-body sensing systems, there has also been increasing interest in using bodily motion as well as gaze behaviour for the same purpose [144,300]. Researchers have for example tried to identify correlations of low-level movement features to affective states, such as the velocity of different body parts [209].

A key requirement for developing computational methods for affect recognition from speech, physical and visual behaviour is the availability of extensive and fully annotated datasets. Such annotation is currently performed manually using video annotation tools, such as Anvil or Elan, according to a specific coding system. One of the most well-known coding systems for facial expressions is the Facial Action Coding System (FACS) [73,74,76] and a similar system has recently been proposed for body actions and postures (BAP) [61]. High-quality manual annotation requires appropriate training of expert coders, making it a cumbersome and costly task. For example, it took Dael et al. on average 15 minutes to code each 2.5 seconds portrayal in the Geneva Multimodal Emotion Portrayals (GEMEP) corpus using the BAP coding system [12]. Moreover, the output is susceptible to subjective interpretation, mistakes and omissions.

While attempts to automate this task for mature coding systems, such as the Facial Action Coding System, have been made [82], the same does not apply to annotating body expression. As the interest in affective body expressions increases, so does the demand for tools and methods to support research in the topic. However, to the best of our knowledge, there is currently no software tool available to annotate affective body expressions automatically.

We aim to fill this gap by presenting AutoBAP, a prototype system that automatically annotates body and eye motion data according to the Body Action and Posture coding scheme using data from wearable sensors. AutoBAP uses hardcoded rules that implement the coding guidelines as well as decision trees trained with machine learning algorithms on data collected in a user study. The decision trees were trained during the system development so that our prototype doesn't have to be trained for new users. Results from a user study demonstrate that our system is able to automatically extract 172 behaviour variables from wearable motion and gaze tracking data with good correspondence to manual annotation.

AutoBAP is the algorithmic layer that sits between a wearable sensing system and a graphical user interface. Figure 22 shows an overview of our prototype.

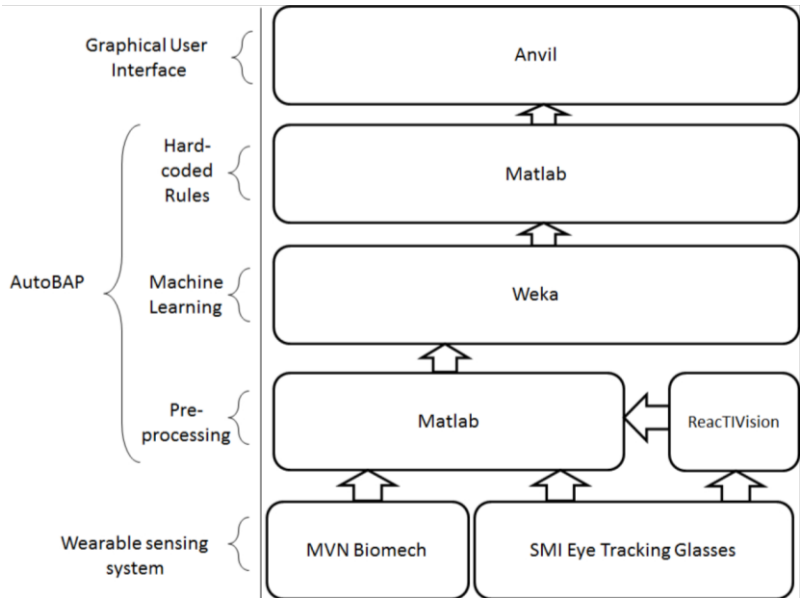


Figure 22 - System Overview. Motion and gaze tracking data are captured with their corresponding tracking system and pre-processed in Matlab. Additionally, we use a computer vision toolkit to track a fiducial marker simulating an interlocutor. We then use the Weka machine learning toolkit to classify the data into initial categories and annotate it using hard-coded rules in Matlab. The system outputs an XML (.anvil) file, which can be visualised and edited in Anvil.

In the bottom layer of our prototype lies the sensing system, which includes a wearable inertial sensors-based motion capture suit and a computer vision-based wearable eye tracker. We opted for wearable solutions as in the future we would like to use our approach to automatically annotate "in-the-wild", i.e. out of the laboratory, behavioural data. In our prototype, we track motion using an Xsens MVN Biomech full body motion tracking system. This is an ambulatory 3D human kinematic measurement system that comprises 17 inertial measurement units (10 in the upper body and 7 in the lower body) and outputs 3D orientation and position of 23 body segments, 22 joints, body centre of mass and raw data from inertial sensors at a sample rate of 120Hz. It transmits its data wirelessly to MVN Studio (version 3.4) which synchronises it to the corresponding frames from an Allied Vision Technologies Prosilica GS650C Ethernet video reference camera (25Hz). Gaze tracking is performed with SMI Eye Tracking Glasses. This is a non-invasive video-based glasses-type binocular eye tracker with automatic parallax compensation at a sample rate of 30Hz, a spatial resolution of 0.1 degrees and a gaze position accuracy of 0.5 degrees over all distances. The glasses are connected with a USB cable to a Windows 8 laptop running the iViewETG software, which streams it wirelessly to another Windows laptop running a custom-built application that records and processes the data.

AutoBAP is the layer above the sensing and is comprised of three components. First, it pre-processes the sensor data. This involves synchronising different sample rates, merging data from different sensors and extracting derived features for the machine learning algorithms. Also, in this prototype we simulated an interlocutor with a fiducial marker placed next to the reference camera and used a computer vision algorithm to extract its position from the eye tracker's scene camera. We do this in the pre-processing stage using the *reacTIVision* toolkit [134]. The second component is the decision tree algorithms. These trees were trained by machine learning algorithms from the *Weka* library [100] using user data. The data collection procedure is described in the next section. The output of the decision trees is then analysed by the third component, which implements the guidelines described in BAP's coding guidelines document and manual. We implemented the rules in *Matlab*.

Finally, after annotating the data, this component down samples it back to the camera's sample rate and creates an XML file with the annotation data. Section 3.5.1 provides an overview of the annotation extraction procedure.

The top layer is the graphical user interface for visualising and editing the extracted annotation. In order to leverage the capabilities and familiar user interface of a widely used tool, we chose *Anvil* [140] for this purpose. In *Anvil*, the user can make any desired changes supported by the platform, such as adding, editing and removing labels (see Figure 23).

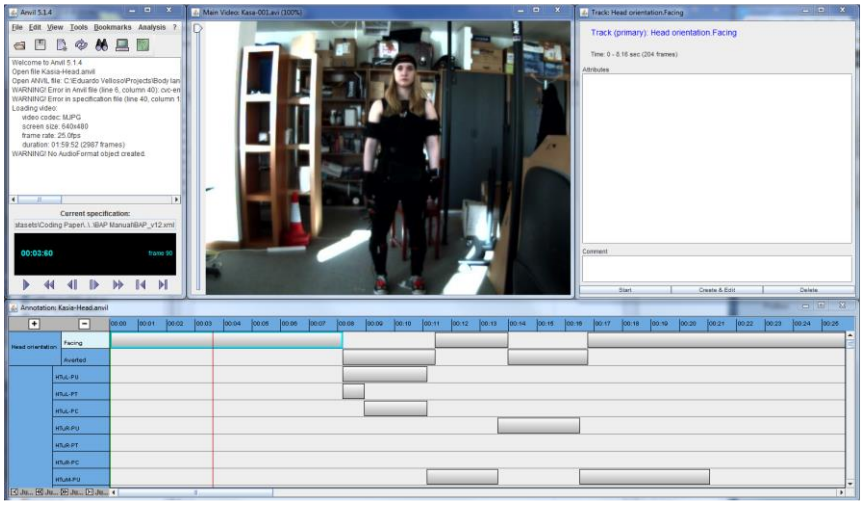


Figure 23 - Anvil user interface. The user can see all labels generated by AutoBAP on a timeline as well as the video recording.

3.5.1 Annotation Extraction

In this section, we describe our approach to classifying actions and postures, which includes decision trees and hardcoded rules. We describe the collection and manual annotation of the data and the selection of features to train the decision trees. We then describe how we adjust the output of these classifiers to adhere to the coding guidelines using hard-coded rules. We exemplify our approach with the example of a right head turn.

3.5.2 Data Collection

STUDY AT A GLANCE

Goal: Train an automatic behaviour classifier.

Method: Data recording

Participants: 4M/2F (18-31y.)

Procedure: Sequence of specific movements to represent all behaviours in BAP

Results: 172/274 labels successfully extracted with a Cohen's Kappa >0.6

The first step in our annotation extraction procedure is to use decision trees training with a machine learning algorithm to classify the data. In order to train our classifier and to evaluate it subsequently, we collected a motion tracking dataset that could cover all labels in the coding system. We collected data from 6 participants, aged 18-31 (mean 24.7), of which 4 were male and 2 were female. They had different body builds, ranging from 1.65m to 1.82m of height (mean 1.74m) and from 59kg to 90kg (mean 73.7kg) of weight. Each data collection session involved only one participant and one researcher and took place at a quiet laboratory environment. The sensing setup consisted of the Xsens MVN Biomech full body motion tracking system and the eye tracker (see Figure 24). Participants stood 2m away

from the camera, which was mounted on a stand 1.12m above the floor, next to a 17" LCD display. Motion and eye tracking data were recorded separately. The average length of the recording for each participant was 9 minutes and 45 seconds.

Figure 24 - Sensing setup. Participants were recorded by a Prosilica video camera (A), positioned next to a fiducial marker (B). They performed actions displayed on an LCD screen (C) whilst being tracked by an SMI eye tracker (D) and an Xsens Biomech motion capture suit (E).

When participants arrived, they filled in a consent form and a personal details questionnaire. We then took measurements of each participant's height, foot size, arm span, ankle height, hip height, hip width, knee height, shoulder width and shoe sole height. These are data that can be input in *MVN Studio* to improve the accuracy of the motion capture. We then assisted each participant in mounting straps with the motion sensors. Each participant was then asked to follow a script of actions so that each behaviour variable in the code appeared at least once in the data. This script was displayed on the LCD screen in a slide presentation showing the instruction and a photo of a person performing the desired posture or action. For example, in the case of the head, participants were instructed to turn left/right and hold the posture; turn left/right without holding the posture and turn left/right repeatedly. The same was done for head lateral and vertical tilts.

We exported the data from *MVN Studio* to a XML file that contained the timestamped position, orientation, velocity and angular velocity of each segment in the global frame and the angle on each joint in their own reference frame. We used the SMI *BeGaze* software to export the data from the SMI glasses to a log file containing the timestamped gaze position in the scene camera reference frame.

3.5.3 Manual Data Annotation

We annotated each recording twice: once to use as input when training the decision trees and once to evaluate the final output of the system. We did not use BAP labels to train our classifiers because some behaviours may be assigned to completely different movements and impact training. For example, the transition phase for head turn towards the lateral middle position might be a left or right movement, as long as they end in the middle. To simplify the training, we annotated the data separately using labels that describe the movement or posture independently of the sequence of behaviours. In the same example, instead of annotating a transition phase to the middle, we annotated a right or left turn accordingly. This way, the machine learning could learn how to classify the direction of the

movement and the orientation of the posture and leave the annotation of what it means in the sequence of behaviours to the hardcoded rules.

We also manually annotated the video recording from the reference camera using *Anvil*, based on the BAP specification file for this platform, which can be obtained from its authors' website. Even though we had no formal training or practice with this particular coding system, since it had just been published, we followed the manual and additional guidelines carefully. We exported the annotation data using *Anvil's* "Export Annotation Frame-by-Frame" feature. This creates a tab-separated text file with a table in which each row represents a frame and each column, a label containing a Boolean value representing the presence or absence of the label in that frame. We then used the timestamps to synchronise the annotation data with the sensor data. All the annotation was performed by the same person. This second annotation was used to evaluate the final output of the system.

3.5.4 Feature Selection

Due to the complexity of human movement, using motion capture data to detect actions and postures also becomes a complex problem. For example, our tracking system can output for each sample up to 794 attributes (4D orientation, 3D position, 3D velocity, 3D acceleration, 3D angular velocity and 3D angular acceleration for each of the 23 segments; 3D acceleration, 3D angular velocity, 3D magnetic field and 4D orientation for each of the 17 sensors; 3D ZXY and 3D XZY angles for each of the 22 joints, the 3D position of the centre of mass and the timestamp), not counting other features that may be derived from those. At a sample rate of 120Hz, this quickly becomes an enormous amount of data, so selecting relevant features increases the speed of training and classification.

Moreover, several behaviours are completely independent of one another. For example, a user may turn his head to any direction independently to his arms configurations. Therefore, classifying behaviours of the head, whilst taking into account the features related to the arm, may improve recognition performance on a training set, but cause erroneous classification on testing data. Therefore, selecting a relevant feature set, also reduces overfitting and increases the recognition performance for further datasets.

Considering that the data were labelled according to the direction of movements and orientation of postures, we treated the annotation of each axis as a 5-class classification problem. For example, in the case of head turns the classes were: right turn, left turn, facing forward, facing to the right and facing to the left. We then used the angular speed to discriminate between movements and the joint angle to discriminate between postures, so that the output of this step would then be used in the subsequent classification procedure.

Some arm postures such as crossed arms, however, involve a specific configuration of more than one axis and segment, so for these cases we used multiple features. Other labels that the coding system is less specific about also require multiple features, but in a less restrictive way. For example, lower limb movements are only coded regarding leg movement or knee bend, so we need to analyse the movement from all segments in either leg to look for movement. For gaze labels we use as features the distance vector between the gaze point and the fiducial marker as extracted by the computer vision algorithm.

3.5.5 Decision Trees and Hardcoded Rules

Extracting a BAP annotation file is a problem that is reduced to filling in a matrix with 274 columns representing each behaviour and one row for each frame in the recording. Each cell in this matrix is a Boolean variable that represents the presence or absence of the behaviour in the frame. A naïve approach would be to train 274 classifiers, but this would

not take into account the relationship between behaviours. Therefore, we reduce the problem even further by grouping sets of exclusive labels. For example, the head cannot be turned to the left and to the right at the same time. Moreover, if the head is moving, it is not, in principle, in any posture (although it can happen, as we discuss in the next session). For each of these sets of behaviours we train a separate classifier using the appropriate feature set, effectively reducing our problem to 28 classifiers. In the case of the head, this leaves us with a separate classifier for head turns, vertical tilts and lateral tilts, with the output being one of five possibilities: the head is either turning to one direction or another, or is being held facing one the middle, one direction or another. We used the J48 decision tree training algorithm available in the Weka machine learning library using the simpler annotation as described previously.

This procedure outputs a table with the predicted labels for each sample in each column, which can be noisy. We smooth the classification output by trying to estimate the onset and offset of each label from the dataset. We use an adaptation of Velloso et al.'s [277] approach to motion modelling to find these points and assign to the segment the mode of the labels it contain, but instead of looking for characteristic points in multiple periodic repetitions of data, we look for these points in a single instance of the movement.

Once we can differentiate directions of movement and orientations of postures, we then take a step back and consider the position of each behaviour in the time series. We hardcoded rules that implement the coding guidelines in the BAP manual. For example, if the segment following a left head turn is a posture held facing forward, we annotate it as a transition phase of a lateral head turn towards a middle position (HTuM-PT), but it is followed by another movement, we label it as an action form sub-unit (Action form.Head-ASU:HTuR). Moreover, if a pattern of repeated action sub-units is detected, we classify it differently (Action form.Head-ASU:HLR, in the case of repeated left and right head turns turning into a head shake). Also, we combine transitions and configurations to extract posture units. Figure 25 exemplifies the classification possibilities for a left head turn.

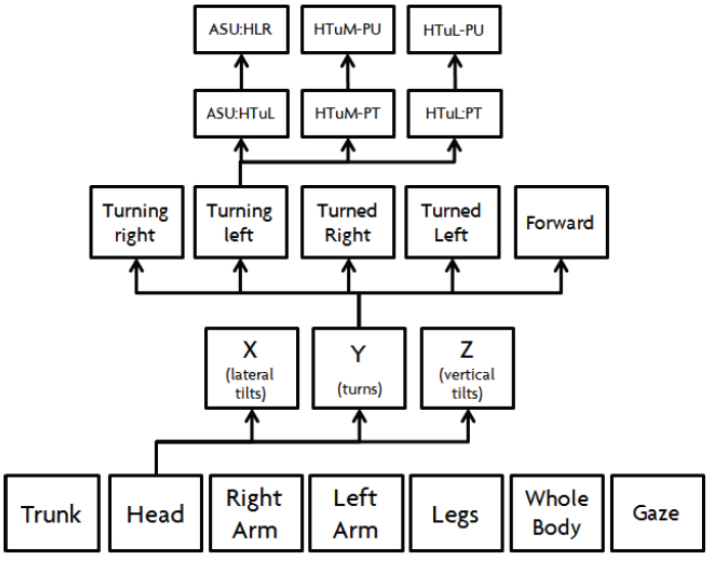


Figure 25 - Example annotation extraction for a left head turn. Input is the data recorded using the motion capture system (A). We then analyse each component of the movement independently (B) and use machine learning to identify the direction of movement or the orientation of the posture (C). Using hard-coded rules that follow BAP's coding guidelines, we analyse the temporal context of the segment and assign the appropriate label (D). Depending on the context, we also combine segments into

parts of larger actions such as head shakes and/or extract other labels such as posture units (E).

3.5.6 Exporting the Annotation

Our classifiers output a matrix in which each column contains a Boolean value for the presence or absence of each label and each row represents a data sample from the motion tracker. We reduce the sample rate of 120Hz to 25Hz in order to match the frame rate of the video recording by taking the 4 or 5 samples corresponding to each frame and creating a data point with the mode of the labels in that interval. We then group intervals with the same label and write it to an XML file according to the Anvil file format. This allows us to visualise and edit the annotation data using Anvil's graphical user interface as if the annotation had been performed manually.

3.5.7 Evaluation

We evaluated the system using cross-validation, using the data from five participants for training and one for testing. We then compared the output of the system with manually annotated data by calculating the agreement between manual and automatic annotations using Cohen's kappa [53] based on the presence or absence of a behaviour unit on each frame in the portrayal. This is a measure of inter-annotator agreement that takes into account the agreement occurring by chance. We considered the labelling successful when the kappa was over 0.6 [84]. Table 3 shows the reliably annotated labels. Posture units shown on the table include all related labels (PU – posture unit, PT - posture transition and PC - posture configuration).

Table 3 - Behaviours annotated with Cohen's Kappa over 0.6

Category	Labels
Head	Facing, Averted, HTuL, HTuR, HTuM, HTiL, HTiR, HTiM, HVU, HVD, HVM
Trunk	Facing, Averted, TLF, TLB, TLMF, TLL, TLR, TLML, TRL, TRR, TRM
Whole Body	BF, BB, BMF, BL, BR, BML
Arms	LA/RA side, LA/RA front, LA/RA back, LH/RH neck, AA crossed, AA front, A hold A front/back, AA sym, AA asym
Gaze	Toward, Upward, Downward, Averted Sideways, Eyes Closed
Head Action Form	HTuL, HTuR, HTiL, HTiR, HVU, HVD, HVUD, HLR
Trunk Action Form	TLF, TLB, TLL, TLR, TRL, TRR, TLLR, TRLR, TLFB
Arms Action Form	AA sym, AA asym, wrist, elbow, shoulder, up, down, forward, backward, left, right, toward, updown, left-right, forward-backward, circular, retraction
Lower Limbs	Knee bend, leg movement

3.6 Limitations

The labels we attempted to extract in this work were limited by the capabilities of the tracking system. In the future, we would like to explore additional sensing modalities to detect the remaining labels, such as hand tracking and touch sensors.

For this work, we recorded a scripted dataset to cover as many labels as possible, but this means that the actions were not natural. Future work will include recording unscripted affective data to improve the training dataset and to evaluate the classifiers in a realistic dataset. This will also allow us to explore the classification of action functions, such as emblems, illustrators and manipulators as well as the possibilities of using automatically extracted labels for affect recognition.

In this first prototype we simulated the interlocutor as a fiducial marker and annotated gaze according to the distance between the gaze point and the fiducial marker as extracted by a computer vision toolkit. In the future, we would like to replace this for a face recognition system, so the system may be used in a real life setting.

In this work, we attempted to annotate as many labels in BAP as possible. However, the coding system was initially created to code the data in a specific dataset, the Geneva Multimodal Emotion Portrayals (GEMEP) corpus, in which actors portray emotions while standing up and being recorded by face and upper body cameras [12]. Therefore, the scope of the coding system is limited to behaviours expressed in such a way. By capturing body expressions with a tracking system, the coding system could be extended in the future to cover more behaviours such as specific leg movements and sitting down postures.

We started the paper by arguing for a combination of the low-level data provided by motion capture systems with high-level posture and action units annotated manually. In the future, we would like to extend the coding specification to leverage this combination. This way, automatically extracted labels could include additional data such as range of motion and average speed for action units and average orientation angle for posture units.

3.7 Conclusion

The Body Action and Posture coding system is still in its early days at the time of writing. As the coding system matures and increases in adoption, studies in Affective Computing will lead to a better understanding of how these labels correlate to affective states. Hence, the automatic extractions of such labels will make it feasible to implement affect recognition systems that take domain knowledge into account.

Our study results show that *AutoBAP* can encode a wide range of BAP units. Some subtle postures were not picked up by our motion capture system despite the fact that we used state-of-the-art motion and gaze tracking systems. For example, the motion capture suit does not include sensors on the fingers, so in this prototype, we did not attempt to label finger actions. Also, even though Xsens's proprietary algorithms extract a very accurate model of motion from the available sensors, the orientation of some segments where no sensors are attached to, such as the neck, are inferred from the orientation of other sensors. This makes the detection of some movements such as neck retractions and extensions more difficult.

In this study, we recorded and annotated very specific and controlled scripted movements, in order to have an unambiguous and comprehensive dataset for training our algorithms. We demonstrated that our approach can classify these datasets accurately but we clearly need to validate our system using other datasets. Also, the training and testing data we used were labelled by the same person. We started from the assumption that the coding system is reliable enough so that two independent raters may end up with a reasonably similar result, as suggested by Dael et al. [61], but we can't make any statements about how the system would perform when compared to third-party annotations. We limited the scope of this study to objective movements, so we did not attempt to classify action functions, such as emblems and illustrators. Due to the wide range of possibilities for such gestures, classifying them becomes a whole challenge on its own.

Our prototype is currently coupled to the chosen sensing system and annotation GUI, but we posit that our approach would be transferable to others. We chose Anvil as the export format as BAP's original specification was published in this format. As it is an XML file and BAP is, in principle, compatible with other annotation tools, it should not be a problem to convert it to other formats. See Schmidt et al. for a description of an effort to convert between annotation formats [234]. Other eye trackers could also be used with few adjustments as long as it provides a scene camera to track the interlocutor or some other means to extract relative orientation and position. Using other motion trackers might be more complicated though. Motion capture systems vary widely in terms of accuracy and which segments they track. While a different implementation would be needed to match the new features to annotation labels, the implementation procedure we described could still be applied.

4 A Survey of Lower Body Interactive Systems

"The human foot is a masterpiece of engineering and a work of art."

Leonardo da Vinci

We now begin to develop our second proposition—that the lower limbs can provide an effective means of interacting with computers beyond assistive technology. Before we began this investigation, our impression was that very little work had been conducted in the area. To our surprise, we found a substantial amount of literature scattered in HCI, Ergonomics, Accessibility, amongst other fields. Because this work had never been surveyed before, we conducted a thorough literature review, which is presented in this chapter.

In December, 1968, Douglas Engelbart delivered what later became known as The Mother of All Demos. In this famous 90-minute presentation, Engelbart introduced the world to the mouse, amongst other prototypes of the fundamental elements of graphical user interfaces (GUI). Whereas it is widely known that Engelbart and his team created the mouse, it is often forgotten that before reaching this design, they explored different prototypes operated by the feet [78,79]. Since then, research in HCI and other fields has given rise to a large variety of computer interfaces operated by the feet, right through to work that employs the feet in mobile and wearable contexts, on interactive floors, and in smart environments. However, in spite of the volume of research conducted on the topic, there is no single reference that systematises the literature. This chapter aims at filling this gap, with a comprehensive review of foot-based interaction.

Work in foot-based interaction emerged from different motivations: the feet provide an alternative to the hands for accessible input [42,249], they can reach areas that are awkward to reach with the hands such as floors [7] and the bottom part of walls [131], they provide a natural mapping to locomotion tasks [70,117], they provide additional input channels for assisting other modalities in complex tasks [91,242], etc.

These different motivations have resulted in the development of a large variety of foot-enabled devices, and research contributions from different communities. For example, the different foot mice and joysticks explored in accessibility research; the variety of sensor-

enabled trainers and insoles created by the wearable computing community; and the diverse ways of tracking the feet unobtrusively with colour and depth cameras that resulted from computer vision research.

This chapter reviews devices and interactions that involve our lower limbs. Considering that when we move our legs we invariably also move our feet, for the sake of simplicity, we refer to these as foot-based interactions. To display the breadth of research that has been conducted, the scope of this survey is broad, covering works that describe foot-operated, foot-worn, and foot-tracking devices; and studies that evaluate interactions afforded by the feet. Also, to put such interactions in context and to give a theoretical background to the understanding of users' capabilities, we review the literature on the anatomy, biomechanics and psychology of the behaviour of the lower limbs.

Regardless of the input or output modalities involved, HCI involves users, systems and the interactions between them. Understanding users helps the design of ergonomically optimal, more widely accessible and culturally compatible interfaces. Understanding systems provides both an awareness of tools available to capture input and provide output, as well as an appreciation of existing systems that provides inspiration and direction for future work. Understanding interactions provides a common vocabulary for the design of interactive systems as well as an awareness of user performance limitations.

This work contributes an analysis of foot-based interactions based on these three lenses. From the user perspective (Section 4.2), we analyse the lower limbs' anatomy and movement, as well as the implications for design created by the pose in which users interact with such systems. We also discuss accessibility and cultural issues. From the system perspective (Section 4.3), we first analyse the different ways of capturing input from the feet—mediated sensing, intrinsic sensing and extrinsic sensing—and how these systems differ in their properties. We also discuss their output and how they provide feedback to users.

Finally, from the interaction perspective (Section 4.4), we analyse four categories of actions that users employ when interacting using their feet: deictic, manipulative, semaphoric and implicit actions. These three perspectives overlap substantially, as one perspective depends on the other to create interactive experiences, but they provide a structure for discussing the most important elements for designing interactions that use the feet as an input modality.

Figure 26 shows our methodology. The survey presented on this chapter, highlights gaps in the literature that we address with a series of empirical studies in the next chapter. These studies informed the design of foot-based interaction techniques that we describe in section 6.2.

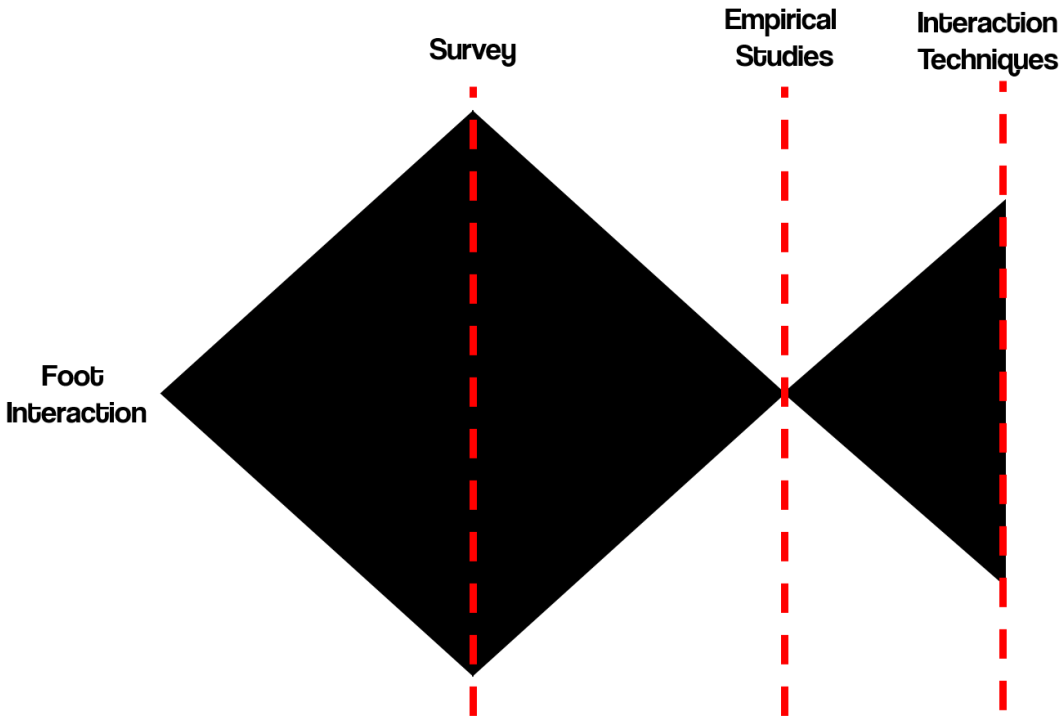


Figure 26 – To explore the domain of foot-based interaction, we conducted a broad survey of foot based interaction (Chapter 4), a series of empirical studies (Chapter 5), and designed novel foot-based interaction techniques (Section 6.2)

4.1 Related Work

We are not the first to attempt to systematise the work on foot interaction. When Pearson and Weiser began the development of their *moles*, they provided a historical classification of the feet in the interaction with devices. In the pre-industrial era, their function was to transmit both power and control (e.g. the horseman's stirrup, the farmer hay fork and shovel, the pipe's organist's bellows and foot keys, the potter's kick wheel). With the advent of electricity and other means of providing power, their function shifted to control alone (e.g. car pedals, arcade games, gas pressure controls, guitar effects pedals). Finally, they were used for foot-mediated input for computers (e.g. flight controls for aircrafts and simulators and volume and sustain controls in music synthesisers) [206]. Whereas this classification puts the role of the feet as an interaction modality in a historical context, it does not provide a structure for modern devices.

Rovers and van Essen's taxonomy classifies foot interactions according to their complexity: (1) simple toggle actions (e.g. foot switches), (2) single parameter (e.g. pedals), (3) multiple parameters (e.g. moles) and (4) intelligent footwear (e.g. Adidas "1") [226]. This classification is also not ideal for modern devices for two reasons. First, because of the miniaturisation and decrease in cost of electronic components, most systems provide multiple channels of input, and hence fall into the third category. Second, there is an overlap between the fourth category and the others, as intelligent footwear may also provide toggle actions and control one or more parameters.

Because of the wide variety of foot-based interfaces found in the literature, rather than trying to find an all-encompassing taxonomy, in this paper, we analyse the literature under three different lenses: the users, the systems and the interactions between them. From the user perspective, we draw from the literature in Biomechanics and Kinesiology [157] to analyse specific movements of the lower limbs. Saffer performs a similar analysis for full body gestures and touch interfaces [228].

From the system perspective we investigate input and output devices. Our classification borrows from general input devices taxonomies, such as Hinckley et al's [111,113]. For certain categories of devices, we refer readers to more specific surveys, for example, on pedals [270] and on locomotion interfaces [117].

Finally, from the interactions perspective, we classify different actions that can be performed with the lower limbs. Karam and schraefel defined a taxonomy for hand gestures in HCI [136] and proposed five categories for gesture styles: deictic, manipulative, semaphoric, gesticulation and language gestures. For an overview of gesture taxonomies, see Billinghurst and Buxton [23].

4.2 Users' Characteristics

The design of interactive systems is usually optimised for the movement and capabilities of the hands. Therefore, it is essential to understand the strengths and limitations of the lower limbs, especially in comparison with the arms and hands, in order to design interfaces that take their motion range, weight and speed into account.

We start our discussion of the user perspective by looking at the anatomy of the legs and feet (Section 4.2.1) and how the movement on each of their joints is used for interaction (Section 4.2.2). We then analyse how the pose of the user (sitting, standing, or walking/running) impacts interaction (Section 4.3.3). Finally, we look at accessibility issues (Section 4.2.4), as well as the nonverbal and cultural issues associated with the behaviour of the feet (Section 4.2.5).

4.2.1 Anatomy

One of the features that sets humans apart from other primates is upright walking, which could date from the earliest phase of human evolution [160]. Whereas our close cousins use all four limbs for locomotion, we only use our legs. This provided an evolutionary advantage as it allowed us to carry more food, to better gather small food from short trees, to expose less skin to direct sunlight, to free our hands to use tools or carry babies over long distances, to spend less energy when walking at reduced speeds, to see further whilst walking and to appear more threatening to predators [282]. The downside is that we lost speed and agility and have a much more reduced ability to climb trees [160]. Also, due to the constant muscle tension applied to stabilise our bodies and to the much shorter length of the toes compared to the fingers, we lost *prehensility*—the ability to grasp—in our feet. The direct implication for HCI is that graspable interfaces, such as the regular mouse, are unsuitable for the feet. Therefore, interfaces with moving parts often need to provide some way of securing itself to the foot, for example, by offering straps or high-friction surfaces.

The foot is a complex structure comprising 26 bones (tarsals, metatarsals and phalanges), over 100 ligaments, muscles and tendons that work together to maintain balance and propel the human body. The bones of the foot are arranged in three arches, two along its length and one across it. These arches stabilise our bodies in the upright position while giving an elastic springiness to it [65]. The foot can be divided into three parts: the hindfoot (where the heel is), the midfoot, and the forefoot (where the ball and toes are). Because of this structure, the feet have a very distinctive and asymmetrical shape that can be recognised by vision-based systems [7]. In the gait cycle, these parts touch the ground in sequence, a pattern that has been explored in several projects to estimate user movement (Section 4.3).

The entire lower extremity of our bodies weigh on average approximately 31.2% of our body mass, of which 19.4% is from the thighs, 9.0% from the legs and 2.8% from our feet [67]. Because the lower limbs weigh a lot more than the upper limbs, their movement tends to be more tiring and lead to cramps if used extensively [78].

4.2.2 Kinematic Analysis of Joints

For our purposes, the movement of the lower limbs is mostly performed by three joints on each leg: the ankle, the knee and the hip. Table 4 shows the distribution of the ranges of motion for each movement of these joints [223].

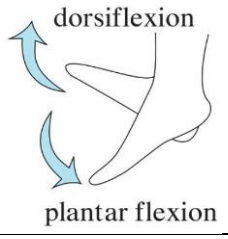
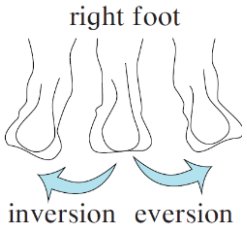
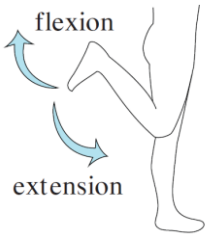
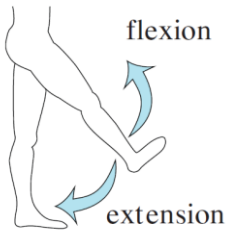
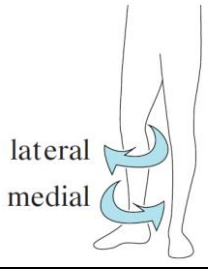
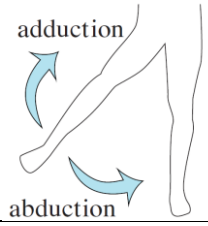
4.2.2.1 Ankle

The ankle joint is capable of three types of rotation, each in two directions: dorsiflexion/plantar flexion, abduction/adduction and inversion/eversion. Dorsiflexion is the movement that decreases the angle between the top of the foot and the leg. Plantar flexion is the movement that increases this same angle. These are the movements used to operate pedals [147]. Momentary pedals, which require users to push against a spring, require more force on the plantar flexion, since the spring assists the dorsiflexion, but rocker pedals require the user to push in both directions. Depending on where the foot is anchored, these movements can be interpreted as two separate gestures. If the foot is anchored at the ball, it is considered as heel tapping and if it is anchored at the heel, it is considered as toe tapping. Pedals typically anchor the foot at the heel, but in English et al.'s control, its vertical movement was anchored at the ball [79].

Inversion is an inward twisting movement, whereas eversion is an outward twisting movement. The range of motion along this axis is very limited. These movements are often combined with other rotations into supination (a triplanar movement in which the foot moves down and towards the centre of the body, combining inversion, plantar flexion and adduction) and pronation (a triplanar movement of the subtalar joint in which the foot moves up and away from the centre of the body, combining eversion, dorsiflexion and abduction). Supination and pronation are the movements typically used to move foot joysticks horizontally as they allow for a shift of weight of the foot with little movement.

Abduction is the movement of the foot away from the centre line of the body and adduction is the movement towards it. As a gesture these movements are interpreted as heel rotations (if pivoting around the heel) or as toe rotations (if pivoting around the toe) [236]. An example of an interface that is controlled by abduction and adduction is Zhong et al.'s *FootMenu* [296], in which the user pivots the foot around the heel to control the horizontal movement of the cursor.

Table 4 - Normal ranges of motion of the lower limbs joints in male subjects, 30-40 years old.

		Range of Motion(°)		
Joint	Movement	Mean	SD	
Ankle	Dorsiflexion	15.3	5.8	
	Plantar Flexion	30.7	7.5	
	Inversion	27.7	6.9	
	Eversion	27.6	4.6	
Knee	Flexion	143.8	6.4	
	Extension	1.6	2.8	
Hip	Flexion	120.3	8.3	
	Extension	9.4	5.3	
	Medial Rotation	32.6	8.2	
	Lateral Rotation	33.6	6.8	
	Abduction	38.8	7.0	
	Adduction	20.5	7.3	

4.2.2.2 Knee

The knee has two degrees of freedom: rotation and flexion/extension. Because knee rotation assists foot abduction/adduction, we will not treat them separately. Knee flexion is the movement that decreases the angle between the leg and the ankle, whereas knee extension is the movement that increases it. As a gesture, these movements combine into a "kick" [193]. Han et al. investigated user accuracy in the direction and speed of kicks [104]. The authors found that targets should cover at least 24° and that users have difficulty in controlling the velocity of the kick, but can remember two broad ranges of velocity.

4.2.2.3 Hip

The hip rotates in three directions: flexion/extension, abduction/adduction and outward/inward rotation. Because rotations around the hip involve moving the whole leg, they are usually tiresome to be used for HCI, but they often assist the movement of other joints. For example, when standing upright, kicks can be enhanced by using the force from the leg. Abduction and adduction are used when moving foot mice horizontally. The hip can also be moved to shift the centre of mass of the body, which is used in pressure sensitive interfaces, such as the Wii Balance board.

4.2.2.4 Toes

Because the toes are harder to control than the fingers and are often covered by shoes, they are very seldom used for interaction. The exceptions are toe switches embedded into shoes, such as the one described by Thorp [265] and the ACHILLE insole, that contains a toe switch for controlling a prosthetic arm [42].

4.2.2.5 Multiple Joints

Different combinations of knee and hip movements allow the foot to move in different topologies. Pearson and Weiser presented four topologies for surface-based foot interaction that illustrate these movements when constrained by a desk well (i.e. the space under the desk) whilst seated (see Figure 27) [206]. A planar topology is defined by a plane that might be tilted by a certain angle. A cylindrical topology ideally has a radius equal to the height of the user's knee, with the main axis crossing the knee. A toroidal topology is defined by a minor radius equal to the height of the knee centred at the knee and a major radius equal to the length of the thigh centred at the hip. Finally a spherical topology has a radius equal to the height of the knee centred at the knee.

These different topologies aim at facilitating the movement for specific joints: in the spherical topology vertical and horizontal movements are optimised for the knee; in the toroidal topology, vertical movement is optimised for the knee and horizontal movement for the hip; and in the cylindrical topology, vertical movement is optimised for the knee. Vertical and horizontal movement in the planar topology and horizontal movement in the cylindrical topology require combinations of knee and hip movements to reach the whole space, but may be simpler for users to understand.

All these topologies constrain the movement of the foot to a two-dimensional plane existing in the three-dimensional space. When this constraint is removed we have the free three-dimensional movement common in movement-based interactions, such as in full body games. These movements use a combination of several joint rotations.

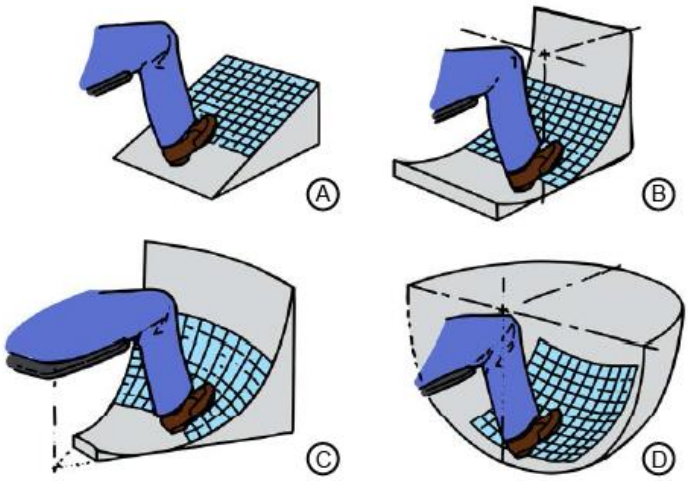


Figure 27 - Topologies for surface-based foot movements in a seated pose (adapted from Pearson and Weiser [206]): (A) Planar; (B) Cylindrical; (C) Toroidal; (D) Spherical

4.2.2.6 Gait

When walking/running, users will use combinations of all these movements. The gait cycle (i.e. the pattern of movement during walking) comprises four stages. It begins when the hindfoot touches the ground (heel strike). Then, the forefoot touches the ground, stabilising the foot and the body (forefoot contact) until the weight of the body is directly over the foot and the opposite foot is swinging from the rear of the body (midstance). Next, the heel lifts from the ground and the weight shifts to the front of the foot, as the opposite foot touches the ground (heel off). Finally, the foot pushes the body forward and enters the swing phase until the cycle restarts (propulsion).

4.2.3 Pose

The pose in which the system will be used significantly affects the design of the interface. We found in the literature three main poses in which foot-operated systems are used: sitting, standing and walking/running. We analysed how the poses affect different properties of the interaction, namely the users' interaction range, the gesture vocabulary, fatigue, challenges for design and operation of other devices (Table 5).

4.2.3.1 Interaction Range

When seated, a user's interaction range is limited to the feet's reach. While swivel chairs allow for reaching locations beyond that by rotating or pushing the chair, chairs with no moving parts require users to either adjust their pose or clumsily reposition the chair when trying to reach further targets. Also, targets may be placed on the chair itself, for example, switches mounted on the legs of the chair. By standing upright, users are able to reach further and by walking towards targets, they can reach indefinitely far targets.

Table 5 - Properties of different poses

Pose		Range	Gesture Vocabulary	Fatigue	Challenges	Other Devices
Sitting		+	+++	++	Desk well, Occlusion	Desktop computers
Standing		++	++	++	Balance, Tracked area	Public displays, multitouch tables, mobile devices
Walking Running		+++	+	+++	Limited attention and cognition	Mobile devices, music players, artistic installations

4.2.3.2 Gesture Vocabulary

The type of contact between feet and floor in a pose determines the range of possible gestures. The sitting pose allows users to take their feet off the floor simultaneously, thus multi-feet and mid-air gestures are possible. Yet, lifting both feet repeatedly or for a prolonged time leads to fatigue. Standing limits the available vocabulary to single-foot gestures, as the other foot maintains the body's balance. At the same time, however, the increased mobility allows for larger gestures, such as kicking or jumping, and for reaching further targets. When walking/running, arbitrary gestures are more difficult because the feet are busy in the gait cycle. Instead, the movement itself is often used as replacement, for example by mapping the real world movement to movement in virtual environments, or by using different walking patterns to issue commands [288].

4.2.3.3 Fatigue

The pose in which the user interacts with the system will also dictate how long the user may interact with it. Users typically have no problem sitting down and to a lesser extent, standing upright, for long periods of time, but walking and running will be tiresome to different degrees depending on the user's physical fitness. Whereas there are no long-term studies on foot interaction in HCI, piano players and car drivers are able to operate pedal-based interfaces for extended periods of time.

4.2.3.4 Challenges

When users are sitting, they are often sat in front of their desks. This spatial configuration constraints the movement in two ways. First, the movement is restricted by the size of the desk well [206]. This not only limits the area where users may move their feet, but because the desk well is often cluttered with cables and power plugs, the movement may be affected. Second, the desk occludes the feet, which prohibits direct input devices, such as *Multitoe* [7].

When users are standing, movement is usually constrained by users' balance, which will determine how well they can perform mid-air or floor-touch gestures in a stable manner. The biggest constraint in walking/running foot interaction is the limited attention and cognition as the user is busy moving through the world. Any deviation of users' normal gait pattern may increase the risk of tripping or losing balance.

4.2.3.5 Interaction with Other Devices

The pose is also influenced by the choice of other devices with which the foot interface will interact. For example, when sitting down, foot interfaces are usually used to interact with desktop computers, together with mice and keyboards. When standing up, they are usually used for interaction with public displays, multitouch tables and mobile devices. When walking/running, the feet normally interact with music players, mobile devices and artistic installations.

4.2.4 Accessibility

Foot-operated interfaces offer an accessible alternative to hand-operated interfaces for people with problems in their hands, including arthritis, carpal tunnel syndrome, limb loss, etc. Several works in the literature investigate these devices explicitly for this purpose [42,249]. While these interfaces may provide relief for tired wrists, continued use may also strain the ankles, causing further pain and discomfort. It can also be more tiring and lead to cramps if used extensively [78].

Interfaces that are exclusively operated by the lower limbs create new accessibility problems. People in wheelchairs and crutches or with other disability or impairment on the lower limbs will have difficulty or even impossibility of using such devices. Also, short people might find difficult reaching far targets on the floor if sitting on a high chair.

4.2.5 Nonverbal Behaviour & Cultural Issues

Both scientific and anecdotal evidence suggest the feet give away clues to our internal states. Joe Navarro, an ex-FBI counter-intelligence officer and body language expert considers the feet as "*the part of the body that is most likely to reveal a person's true intentions*"[183]. In social interactions, we tend to focus on each other's faces, so the legs and feet tend to escape our attention. This makes the lower limbs particularly good at providing clues to how people are really feeling—what psychologists call nonverbal leakage [179].

In Section 4.2.1 we explained how our lower limbs evolved to support our bipedalism. As they became our main means of locomotion, our limbic system—the part of our brains responsible for, amongst other functions, our emotions and our fight-or-flight mechanisms—evolved to quickly activate the legs and feet to escape from danger or confront predators. Even though we do not face such challenges today, these hardwired evolutionary mechanisms still manifest themselves in our nonverbal behaviour. For example, quick movements of the feet are indicative of anxiety. In a study with students learning foreign languages, Gregersen reported that anxious students continuously bounced, jiggled and tapped that feet, whereas non-anxious students only crossed and uncrossed their legs a few times [93].

Not only feet movements send nonverbal signals, but also the overall posture. For example, Mehrabian [171] relates the symmetry of leg posture to how relaxed the person is: the more asymmetrical the posture, the more relaxed the person. He defines four categories of symmetry in ascending degree of relaxation: symmetrical position of the legs with both feet flat on the floor and the insteps touching; symmetrical position of the legs with both feet flat on the floor and the insteps not touching; asymmetrical stance of the legs with both feet resting flat on the floor, asymmetrical stance of the legs with one or both feet partially lifted off the floor.

The behaviour of our feet is also unconsciously influenced by the behaviour of the people around us—because of the so-called *Chameleon Effect*, we tend to mimic the behaviour of the people we are talking to. In Chartrand and Bargh's study, participants were more likely to tap their feet during a task when their confederate was also tapping his feet [48]. Different leg postures also influence the rapport between people interacting. Harrigan et al. investigated nonverbal cues in physician-patient rapport and found that high rapport physicians were more likely to sit with their legs uncrossed, with their bodies orientated toward the patient, but there was no difference in feet movement between high-rapport and low-rapport doctors [105].

Certain behaviours of the legs and feet vary in different cultures. In several cultures, especially in South Asia, the feet are considered dirty and pointing with the feet or exposing the sole of the foot may be considered rude or insulting [132]. Wagner et al. investigated the social acceptability of touching on-body targets and found that targets on the lower limbs were significantly less acceptable than on the upper limbs [280]. In China, bound feet used to be considered sexually appealing in women and some men preferred never to see their feet, which were constantly concealed by shoes and wrapping. This practice was banned in 1949 and the ban remains in effect since.

Some leg postures can also be perceived differently by distinct cultures. For example, certain American men can perceive the way European men cross their legs (with one knee crossed over the other) as slightly effeminate [179]. However, with more cultural exchange around the world these cultural differences tend to be minimised as we can see both Europeans sitting with the ankle over the knee and Americans with one knee over the other [179].

These are just a few examples of how people's psychological states and culture are perceived and manifested through their feet. For an in-depth treatment, see Morris [179] and Argyle [6]. We return to the topic of how this natural behaviour can be leveraged in HCI in section 4.4.3.

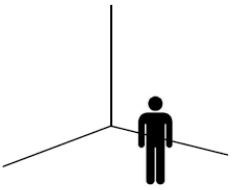
4.3 Foot-Based Systems

In the previous section we discussed how users' body pose and movement affect the interaction. In this section we describe and categorise research prototypes and commercial systems that take input from the feet (Section 4.3.1) and discuss different ways of providing feedback in foot-based interactions (Section 4.3.2).

4.3.1 Input Sensing

Foot-operated input devices exist in a variety of shapes and sizes—from small foot mice, to room-sized augmented floors. In this section, we classify these devices into a taxonomy according to how they capture input from the feet: *mediated*, *intrinsic* and *extrinsic* sensing (see Table 6). Mediated sensing happens when the feet are not tracked directly, but rather through devices operated by them. Intrinsic sensing refers to when the feet are tracked through sensors directly attached to them and extrinsic sensing refers to when the feet are tracked through sensors placed on the environments.

Table 6 - Categories of foot input sensing

Category		Instances	Passive Feedback	"Always-On"
Mediated		Foot switch, pedal, knee control, foot mouse, foot joystick, trackball, moles, balance boards	+++	+
Intrinsic		Augmented shoes, augmented insoles	++	+++
Extrinsic		IR-based trackers, vision- and depth-based trackers, laser range finder, augmented floors and surfaces	+	++

4.3.1.1 Mediated Sensing

In mediated sensing, instead of tracking the feet directly, sensors track devices operated by the feet. This category comprises mechanical devices such as foot mice, trackballs and pedals, which contain moving mechanical parts that capture input from the feet. As a result, such devices provide immediate passive haptic feedback on the actual input action. They are usually found in the form of computer peripherals, and hence only capture input when the user is directly interacting with the device.

The oldest, most widely known and most thoroughly studied foot interface is the **pedal**. Pedals are employed in a wide variety of contexts: cars, bicycles, boats, aircrafts, pianos, harps and guitar effects are some examples from outside the world of computing. Due to the safety-critical role that pedals play in cars and machinery operation, studies looked into finding optimal pedal designs, which go as far back as 1942 [13]. See Trombley [270] for a review of early work on pedal operation and Rosenblum [224] for a review of the role of pedals in pianos.

The simplest form of a pedal is a binary **switch**. Examples include foot operated light and tap switches and transcription pedals that have multiple switches for controlling playback. As with any switch, they can be latching or momentary, depending on whether it returns to its initial state once the user releases it. As Sellen et al. point out, momentary pedals provide advantages for selecting the mode of operation of software systems because they require users to actively maintain the state by holding down the foot—its kinaesthetic feedback is more difficult to ignore reminding users of which mode they are currently in [238].

Rather than sensing discrete states, pedals also allow for controlling continuous parameters, such as the acceleration of a car. In this mode, it is necessary to choose an adequate mapping between the pedal and the parameter being controlled: a 0-order control alters the value of the parameter directly, while a 1st-order control alters the rate of change of the parameter. For example, Kim et al. compared 0-order and 1st-order pedals for setting font sizes in a text entry task and found the performance of the 1st-order control to be comparable to the mouse [139]. Similarly to foot switches, continuous pedals can also be momentary or latching.

Other examples of works that use pedals include controlling a 3D modelling application [10], text entry [66], supporting gaze input [91] and toggling the mode of operation of a piano keyboard [176]. Zhong et al. implemented a pivoting pedal that rotates around the heel in addition to up and down [296].

Safety-critical applications of pedals should also take into account some problematic issues. First, people seldom scrutinise the floor when they are working, so they might trip on pedals. Also, "riding the pedal"—which happens when the user stops pressing the pedal, but remains with the foot on top of it—is the most prevalent cause of accidental activation [14].

Despite the pedal being around for a long time, it was not the first foot operated interface for a computing system. Among the first alternatives for controlling a cursor, the same team that invented the mouse developed a knee control consisting of two potentiometers linked to a **knee lever**, which was controlled by the user by pushing the lever side-to-side or up and down [79].

The term **foot mouse** has been used for different kinds of foot-operated devices. In this chapter, we restrict the term to devices that work in the same way as the hand mouse, while being moved by the foot. Therefore, the physical property that is used as input is the position of the foot. These include commercial products such as the *BiLiPro Foottime Foot Mouse* (2006) and research prototypes such as the puck on a Wacom digitising tablet that Balakrishnan et al. used to control the camera in a 3D modelling application [10].

Foot joysticks (also often called foot mice in the literature, e.g. [249]) are controls in which the user nudges the device in a specific direction to control the speed of movement of the cursor in that direction. The *Versatron Foot Mouse* (1984) is the first example of such interface and was controlled by sliding with the foot a rubber platform spring-loaded to return to the central position. The more recent *No Hands Mouse* (2009) uses two devices: one foot joystick to control cursor position and one to control mouse clicks. Such devices are usually mapped as 1st-order controls, similarly to hand-operated joysticks. Garcia et al. studied learning effects of users interacting with this device [88,89]. Research prototypes of foot joysticks include the works of Springer and Siebes [249] and Göbel et al. [91].

Large trackballs have also been used for foot-operated cursor control, such as the *BIGtrack Trackball* and the *AbleTrack Trackball*. Mouse clicks are usually operated by external foot switches. Pakkanen et al. suggest that such interfaces are appropriate for non-accurate tasks [195].

Pressure-sensitive boards use the distribution of the user's weight to control two-dimensional variables. An early video-game controller that used this principle was the *Amiga Joyboard* (1982), which contained the four directional latches of a joystick on the bottom of the board and by leaning in a certain direction, the user engaged these latches and controlled the game. The *Nintendo Wii Balance Board* (2007) uses four pressure sensors to measure the user's centre of balance and has been used for several purposes beyond entertainment including fitness, navigating maps [235] and navigating 3D environments [287].

Pearson and Weiser created several early prototypes of foot interfaces, dubbed "**moles**" , as the "*beasts are situated under-foot*" [206]. These authors proposed the term for the category

of devices that are operated by the feet in a similar manner to the mouse, but it was not picked up by other authors. Instead, the term ended up referring to the multiple prototypes they built in the 1980's that used a rig under the desk to simulate mouse input. In a second work, they implemented a planar mole featuring a pedal that slid in all four directions constrained to a plane [207]. In a third one, they built two versions of the swing mole, in which the cursor is controlled by the right foot, which slid left and right on a platform that rotated along an axis inside the desk well from the front desk edge [208].

A final category of mediated sensing input devices are **locomotion interfaces** for virtual reality. These devices use repetitive movement of the user's limbs to navigate through virtual environments. For a full treatment of this category, we refer the reader to Hollerbach's survey [117]. Hollerbach categorises locomotion interfaces as pedalling devices (e.g. bicycle simulator [35] and the *Sarcos Uniport* (1994)), as walking-in-place devices (e.g. *Gaiter*[264]), as foot platforms (e.g. *Sarcos Biport, GaitMaster* [126]) and as treadmills (e.g. the *Sarcos Treadport*, the *Omni-Directional Treadmill* [64]).

4.3.1.2 Intrinsic Sensing

Intrinsic sensing devices contain sensors directly coupled to the feet. These systems are typically wearable and self-contained, requiring little to no instrumentation on the environment, thus allowing users to move freely. Because of this increased mobility, these systems typically monitor users' walking patterns to make inferences about the user. Such systems usually come in the form of sensors and actuators augmenting users' insoles or their whole shoes. They are typically *always-on*, meaning that they continuously track users as long as they are wearing the device, without explicit user interference.

The first documented wearable computer was in fact manipulated by the feet. Thorp describes how in the 1950's and 1960's he developed a wearable computer to predict the outcome of casino roulette wheels operated inconspicuously using a toe switch in his shoes [265]. The increased interest in Wearable Computing in the late nineties sprung a variety of projects interested in augmenting users' shoes. In an early article on the topic, Mann mentions building trainers that measured his pace [166].

Wearable interfaces rely on sensors and devices worn by the user that capture information from the user's feet. Table 7 shows the sensors used in previous work. The most common sensors in smart shoes are pressure sensors in the form of force sensitive resistors. By distributing such sensors on different points of the sole, it is possible to calculate the weight distribution of the user and infer gait patterns. Bend sensors work in a similar manner, by changing their resistance as they are flexed. These are usually installed in the middle of the foot to detect when the foot is bending, such as in the beginning and end of the gait cycle.

Table 7 - Sensors in selected prototypes of augmented shoes

Reference	Pressure	Acceleration	Gyroscope	Bend	Temperature	Humidity	Light	Distance	Application
In-Shoe Multisensory Data Acquisition System [178]	×				×	×			Patient monitoring
Expressive Footwear [198]	×	×	×	×				×	Artistic performance
CyberBoots [50]	×								Artistic performance
Shoe-Mouse [290]	×	×	×	×					Cursor control
Shoe Shaped Interface [281]	×								Inducing a gait cycle
Intelligent Shoes [120]	×	×	×	×					User identification
ACHILLE [42]	×								Control of prosthetics
CabBoots [86]		×	×				×	×	Assisting navigation
Shoe-Shaped I/O *Interface* [110]		×							Artistic performance
Sonic Shoes [155]	×								Auditory feedback
Rhythm 'n' Shoes [196]	×								Artistic performance
Shoe-Keyboard [262]	×	×							Character input
ShoeSoleSense [168]	×								Virtual Reality

Another widely employed category of sensors are inertial measurement units, which comprise different combinations of accelerometers, gyroscopes and magnetometers. With such sensors, it is possible to use movement information to estimate the position and orientation of the foot, which, in turn, can be used for analysing gait and for explicit interaction with computers.

All the sensors described so far measure the movement, orientation and configuration of the foot, but sensors that attempt to make sense of the environment around it—both inside and outside the shoes—have also been explored in the literature, such as humidity, light and distance sensors.

A challenge for wearable devices is how to power all these sensors and processing units. When we walk, we generate kinetic energy, which can be harnessed by augmented shoes

using piezoelectric elements. These components convert the mechanical stress created by the foot when pushing down against the floor into electric current. Even though the current generated is not much, it is enough to power active RFID tags [191] or low-power components.

Paradiso et al. describe several iterations of trainers—dubbed *Expressive Footwear*—augmented with pressure, bend, position, acceleration and rotation sensors for interactive dance performances [198,199,200,201,202] and interactive therapy [203]. *CyberBoots* was another early work aimed at interactive performances, which consisted of boots that could be worn over the user's shoes to detect walking and leaning patterns from pressure sensors mounted on the insoles [50]. More recently, other works that used foot mounted sensors for artistic purposes include geta clogs augmented with a *Nintendo Wiimote* and a pico projector for guitar performances [110] and sandals that detect foot tapping to create accompanying drums [196].

By installing sensors on users' shoes, it is possible to unobtrusively monitor their gait patterns anywhere they go. Applications for this kind of technology include detecting abnormal gait patterns [49], identifying users by their gait [120], adjusting music tempo to match the user's [114]; navigating in virtual environments [168]; inducing a specific walking cycle [281]; assisting navigation [86]; and even changing the noise of users' footsteps [155]. A commercial example is the *Nike+iPod* (2006) line of trainers, which measures and records the distance and pace of a walk or run. From users' gait pattern it is also possible to infer their trajectories using a technique called pedestrian dead reckoning. Fischer et al. describe how to achieve this using inertial sensors mounted on the foot [82].

Augmented shoes have medical applications. Morley et al. installed pressure, temperature and humidity sensors to measure environmental conditions inside diabetic patients' shoes [178]. The commercial product *SurroSense Rx System* also aims at assisting diabetes patients by collecting pressure data on the insole to help prevent foot ulcers. Carrozza et al. installed switches on the insole of a shoe to control a prosthetic biomechatronic hand [42].

Motion sensors on the feet have also be used to detect gestures for explicit human-computer interaction, including using augmented shoes to emulate a conventional mouse [290] and keyboard [262]. Gesture-based foot interaction is particularly interesting for mobile devices. Crossan et al. investigated foot tapping for interacting with a mobile phone without taking it out of the pocket using an accelerometer on the top of the user's feet [56]. Scott et al. investigated such gestures can be recognized using the sensors in a mobile phone inside the user's pockets [236]. Alexander et al. collected a foot gesture set suggested by users to control a mobile device [2].

4.3.1.3 Extrinsic Sensing

Interfaces in this category rely on sensors placed on the environment to capture data from users' feet from the outside. They usually require little or no instrumentation on the user and are always-on as long as the user is within the tracked area. They typically offer little to no passive haptic feedback.

Several studies of foot interaction rely on **passive infra-red motion capture** systems to track the feet. These systems use infra-red cameras to capture the light reflected by markers attached to different points on the user's legs and feet. They are usually very accurate, but require several cameras depending on how large is the volume being tracked and require direct line of sight between the markers and the cameras. Moreover, the need for special markers attached on the users' bodies makes them unsuitable for casual interactions. *The Fantastic Phantom Slipper* was a pair of slippers with reflective markers and vibration motors used for interacting with a virtual environment [150]. Quek et al. used a passive IR system to track users' feet in order to estimate participants' attentional focus from their feet posture [215].

Vision-based systems use data from one or more colour cameras to extract the feet's position and orientation. The advantage of such systems is that they require little more than a camera, making them easy to deploy. The downside usually comes in loss of accuracy and the need for direct line of sight. *AR-Soccer* is a football game using the camera in a PDA by extracting the contour of the user's foot and detecting its collision with a virtual ball in order to kick the ball towards the goal [193]. Similarly, ur Réhman et al. tracked the feet using a mobile phone using template matching [219]. The *Visual Keyboard* used a vision-based approach to extract users' feet in order to play a musical keyboard [129]. Vision methods have also been used to activate virtual pedals [239,292], predict driver behaviour [269] and control a first-person game [287].

Commercial **depth cameras** such as the *Microsoft Kinect* and the *Asus Xtion* made it easier to track the feet accurately in three dimensions. These cameras are usually cheap and fairly accurate, but they also require direct line of sight with the user. Han et al. used this approach to investigate kick gestures for mobile interaction [104]. At the time of writing, the available SDK of these systems is only able to track legs and feet when the users' whole body is in the field of view. Some works aimed at extending this functionality to be able to track the feet when the rest of the body is not visible. Hu et al. proposed a method to accurately extract a 3D skeleton of user's legs and feet with a Kinect mounted on a walker for the elderly (a.k.a. a Zimmer frame), but it does not run in real time [119]. *Bootstrapper* recognizes users around a multitouch table by looking at their feet using Kinect sensors mounted on the table edges facing down [222]. In the next chapter, we describe the foot tracker we developed based on a Kinect sensor mounted under a desk [242]. Another similar approach is to use a laser range finder. Huber used such system to estimate spatial interest at public displays [122].

As the feet are most of the time in contact with the floor, feet tracking lends itself to using **augmented floors**, which can be implemented with a variety of sensors. The first few prototypes of interactive floors were built for dance performances: *Magic Carpet* used a grid of piezoelectric wires [197], *LiteFoot* used a matrix of optical proximity sensors [95] and Z-Tiles used a modular architecture of hexagonal pressure-sensitive tiles [169,221]. Lopes et al. used a Dynamic Time Warping algorithm to classify foot gestures on a wooden board from audio data [159]. This approach is simple to deploy, but can only detect gestures rather than positions or orientations and may suffer from interference from other sources of noise. Other approaches for augmenting floors include pressure sensitive [192,230] and capacitive floors [128].

Camera-based floors sense users' feet positions with computer vision techniques. *iGameFloor* used four webcams to track users from under a semi-transparent floor [96]. *Multitoe* is a high-resolution Frustrated Total Internal Reflection (FTIR) back-projected floor that allows for precise touch input [7]. *GravitySpace* used the same technology to reconstruct users' poses in 3D above the floor from pressure imprints [31]. *Kickables*, in turn, used *Multitoe* to track tangibles that users manipulate with their feet [233]. Whereas these approaches are highly accurate and provide output on the same surface, they do not scale very well and are difficult to deploy, as they require a lot of changes in the infrastructure. An alternative would be to track users from above, such as in *iFloor*, but at the price of losing tracking accuracy [148]. A variety of companies sell top-projected interactive floors, including *EyeClick*, *Luminvision* and *GestureTek* and such installations have been deployed in shopping malls and other public spaces all around the world.

The feet have also been used to interact with **vertical surfaces**. Jota et al. implemented interaction techniques for interacting with the bottom part of vertical displays, where the hands would not be able to typically reach [131].

Because augmented environments and surfaces are often able to extract the position and orientation of the whole foot in two or three dimensions, systems can use this information in different ways: as a blob, as a hotspot or as relative motion. A **blob** is a set of points that

pertain to the foot. When in three dimensions, this is often called a point cloud. For example, FTIR-enabled floors see a 2D blob, whereas depth cameras see 3D point clouds. Using the feet as blobs effectively increases the size of the target, because any part of the foot can activate it. Due to the relatively large size of the feet, this is more prone to accidental activation, so targets should be large and well spread apart. Examples of works that track the feet as blobs include Paelke et al. and ur Rehman et al.[193,219].

Instead of using the whole blob, it is also possible to reduce it to a single or multiple **hotspots**. This allows for more precision in the interaction. For example, Simeone et al. [242] convert a point cloud to two points in 3D representing the tip of the foot and the ankle and Augsten et al. [7] convert it to a single point in 2D. Augsten et al. also investigated which positions users find intuitive for this hotspot. Their results indicate that there is no universal position agreed by all users, so in their system they implemented a calibration procedure that allows users to customise the position of the hotspot.

The final approach is to ignore the absolute position and orientation of the feet and only take into account their **relative movement**. This is often used for gesture recognition, such as in Lopes et al. [159].

4.3.2 Output & Feedback

In order to close the interaction loop, interactive systems must provide some kind of feedback or output. Types of output in foot-operated interactive systems include visual, auditory, haptic and thermal feedback.

Visual feedback is the primary feedback modality for traditional computing systems, so it is of no surprise that a wide variety of foot-operated systems provide some kind of visual feedback. An important issue for visual feedback is the distinction between direct and indirect input devices. Direct input devices (e.g. a touch-enabled screen) have "a unified input and display surface", whereas indirect devices (e.g. the mouse) do not "provide input on the same physical space as the output" [113].

Direct input can be implemented by using touch sensitive floor displays, such as *Multitoe* [7] or by overlaying the interface with the foot, either through Augmented Reality [193] or by projecting the interface on the floor [170]. This presents a challenge because the feet significantly occludes the screen. Moreover, in order to visualise the output, the user must look down, which can be tiring after extended periods of time. In indirect input, the feet and the display are separate. This creates the need for some representation of the feet on the screen, such as a cursor or other parameter that the feet are controlling. Because the input and output devices are separated, special attention must be paid to the mapping between the two, as incompatible mappings may be cumbersome for the user [45].

Audio feedback is usually implemented in mobile systems, in order to reduce the cognitive visual overload. Because of the natural rhythmic pattern of our gait, several systems used input from the feet to modulate music, especially tempo, when the user is walking or jogging [22,175]. Another application area in which the feet are used to generate audio output is in artistic performances, by mapping dancers' [202] and musicians' [196] feet movements to music parameters. Stienstra et al. explored how auditory feedback can improve speed skaters' performance by sonifying the data captured from force and acceleration sensors on the skates [256].

Haptic feedback consists of forces or vibrations applied on the users' skin to stimulate their sense of touch. This has been implemented in a wide variety of augmented shoes in the form of vibration motors [168,281]. Rovers and van Essen presented an investigation and design guidelines for using haptic feedback at different points on the foot sole [225,226]. These authors' findings suggest that users understand better vibration patterns in the longitudinal direction than in the transversal direction; users recognise static patterns in the transversal

direction, but are confused by moving patterns; and recognising more complex patterns such as a "zigzag" pattern is difficult. *CabBoots* [86] provides height/tilt actuation to help the user to navigate through an environments by using actuators inside the boots to create an angulation on the shoe sole and hence steer the user in the correct direction. A commercial example of mechanical actuation inside the shoe are the *Adidas 1* (2005) trainers, which contained a motor in the middle of the sole that changed the compression characteristics of the heel pad. The *Vectrasense Raven Thinkshoe* (2004) did something similar but with an air bladder in the sole. One commercial example of haptic feedback on the feet for navigation is the *Lechal Shoe*, which vibrates to guide the user on the path he set on his smartphone.

Thermal feedback can be produced by Peltier elements, which create a temperature differential on each of its sides. Matthies et al. included one in an insole and suggested that a rising temperature could be used to make the user unconsciously uncomfortable in situations such as to indicate the player is wounded or bleeding in a game or feedback on how many missed calls or unread messages the user has received [168]. They stress, however, that the human body tends to acclimate to thermal discomfort, becoming less responsive to a constant stimulus. The authors suggest alternating between hot and cold to maintain the sensory response.

4.4 Foot-Based Interactions

So far, we have analysed previous work in terms of users and systems, relating how different body poses and movements impact the design of systems that capture input from the feet and provide some kind of output. In this section, we deal with the dialogue between users and systems. More specifically, we are concerned with the different actions users perform with their feet for interaction.

Karam and schraefel defined a taxonomy for hand gestures in HCI [136] and proposed five categories for gesture styles: *deictic* (gestures involving pointing), *manipulative* ("whose intended purpose is to control some entity by applying a tight relationship between the actual movements of the gesturing hand/arm with the entity being manipulated" [216]), *semaphoric* ("any gesturing system that employs a stylized dictionary of static or dynamic (...) gestures" [216]), *gesticulation* (gestures that accompany speech) and *language gestures* (such as sign language).

We distinguish four categories of feet actions in HCI: *semaphoric* (Section 4.4.1), *deictic* (Section 4.4.2), *manipulative* (Section 4.4.2) and *implicit* (Section 4.4.3). We use the same definitions as Karam and schraefel for semaphoric, deictic and manipulative. Implicit actions comprise the nonverbal behaviour of the legs and feet as well as non-communicative actions, such as walking.

4.4.1 Semaphoric

Semaphoric actions are specific gestures belonging to a dictionary. We compiled a comprehensive list of gestures explored in the literature in Table 8. In Section 3.2 we showed how gestures derive from the degrees of freedom of the lower limbs. In this section we look at how these movements and their combinations are used in interactive systems.

Touch-sensitive surfaces, inertial motion sensors and depth cameras made multitouch and mid-air gestures a reality for consumers [228]. By adapting these technologies for the feet—e.g. multitouch floors [7], IMUs on the feet [104] and depth cameras under the desk [242] researchers have been incrementing the feet gesture vocabulary as well as understanding what are the appropriate mappings between gesture and functionality.

In a guessability study, Alexander et al. investigated users' intuitive mappings between feet gestures and controls for mobile devices, compiling a gesture set with corresponding mappings phone and media control as well as map and browser navigation [2]. They identified two large sets of gestures: discrete and continuous. Furthermore, the authors compared four techniques for implementing continuous mappings for panning actions: displacement-based, rate-based hold, rate-based continuous and flick. Similarly to Kim and Baber's findings for pedal operation [139], Alexander et al. found that users tend to prefer rate-based approaches.

Table 8 shows the gestures that emerged from our literature review. **Toe tapping** is the most common gesture across papers. We attribute its popularity to its low effort, to its historical operation in pedals and to it being analogous to a finger touch. A variation is the **heel tap**. The disadvantage of the heel tap is that it requires users to lift the weight of the leg if sitting or the whole body if standing, which can be demanding on the calf muscle. These gestures are often mapped to selecting an option or activating a certain functionality.

In **toe and heel rotation**, the user pivots the foot with an abduction or adduction movement at the ankle. When performing this action, the foot remains anchored at the heel or toe, so it is usually more comfortable than the swipe, which involves moving the whole foot in a certain direction, requiring some leg effort.

Toe and heel clicking gestures require users to use both feet at the same time. Clicking the heels in Western culture is often associated to film "The Wizard of Oz" (1939), in which Dorothy, the main character, clicks her heels to go back home. For this reason, heel clicking has been used to invoke the system (the interface's home) by Laviola et al. [154].

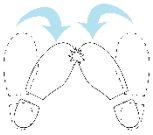
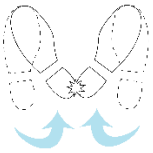
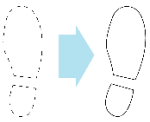
Shaking the foot involves an irregular movement across multiple movement axes. The dismissive semantics associated with the gesture led participants in Alexander et al.'s study to map it to ignore an incoming call [2]. Because it is an easy gesture, with a very distinctive pattern—unlike tapping, for example, which can be mistaken with natural walking—it can be used as a *Whack Gesture*, i.e. inexact and inattentive interaction techniques that require minimum cognitive processing from the user [123].

As immortalised by Daniel Day-Lewis's performance of the real story of Irish painter Christy Brown in *My Left Foot* (1989), some people are able to paint with their feet. Whereas this may sound incredibly hard for most users, it is still possible for them to **trace basic shapes** with their feet. In Alexander et al.'s study, the shape being traced was a circle, but it would also be possible to devise interaction techniques that use other shapes, such as a square or a triangle [2].

A foot gesture that has been particularly well studied is the **kick**. Han et al. investigated how well users can control the direction and velocity of their kicks in selecting radial targets [104]. The authors recommend using at most 5 targets with an angular width of 24o. In Jota et al. [131], kicks are used to transfer objects on a vertical screen between the foot-operated area and the hand-operated area. Schmidt et al. proposed tangibles for feet that users push and kick across a large interaction surface [233].

Stepping is the basic unit of the *Walking-in-place* (WIP) interaction technique [245]. The technique is commonly used for walking around virtual environments, as the proprioceptive information from the body movements makes the user feel more immersed in it. This technique has been used extensively in the Virtual Reality (VR) community and implemented in a wide variety of ways, including sensing in the head-mounted display [245], the Wii Balance Board [284] and the Kinect depth camera [295]. Because this technique is often used for VR locomotion, controlling speed and direction is critical. Among the solutions are the *Gait-Understanding-Driven* (GUD-WIP) [283], *Low-Latency, Continuous Motion* (LLCM-WIP) [81], and *Speed-Amplitude- Supporting* (SAS-WIP) [38] walking-in-place techniques.

Table 8 - Dictionary of Semaphoric Foot Gestures

Gesture	Name	Description	e.g.
	Toe Tap	User raises and lowers the toes	[56]
	Heel Tap	User raises and lowers the heel	[262]
	Toe Rotation	User pivots the foot around the toes	[236]
	Heel Rotation	User pivots the foot around the heel	[236]
	Toe Click	User touches both toes together	[154]
	Heel Click	User touches both heels together	[154]
	Swipe	User slides the foot in a given direction	[154]
	Shake	User moves the foot with short, quick, irregular vibratory movements	[154]
	Shape Trace	User draws the outline of a shape with the toes	[2]
	Kick	User vigorously moves the foot forward	[104]
	Step	User puts one foot in front of the other as if walking	[70]

The gestures described so far sprung mostly from research prototypes, but are starting to appear in commercial products. For example, the *Lechal* footwear line uses foot gestures to tag locations, set destinations, and control navigation.

4.4.2 Deictic & Manipulative

Deictic actions are commonly understood as pointing gestures. When we use our foot to tap on a target on an augmented floor [7] or move a trackball over a target on a GUI [195], we are using deictic actions. Manipulative actions map elements of the physical configuration of the foot, such as position or orientation to properties of system objects. When we drag a foot to rotate a virtual cube [242], push a tangible object across the floor [233] or move a foot mouse to control the orientation of a scene camera [10], we are using manipulative actions. As both types of action involves getting the foot from an original configuration to a target configuration, although for different purposes, we examine them together. More specifically, in this section we review the literature on the movement times, accuracy, reaction times and learning effects of such actions.

4.4.2.1 Movement Times and Accuracy.

The most widely adopted model of human movement is Fitts's Law, which was originally developed to predict movement times for hand pointing, but has been proved applicable in a wide variety of situations [83]. Its most common form is the Shannon formulation, proposed by MacKenzie [164], where MT is the time in seconds to reach the target, D is the distance to the target and W is the width of the target:

$$MT = a + b \times \log_2\left(\frac{D}{W} + 1\right)$$

Early work on measuring movement times for the feet was concerned with pedal operation—for a review, see Kroemer [147]. In this context, Drury applied Fitts's Law to find optimal pedal positions [71]. He proposed a variation of Fitts's Law that takes into account the width of the user's shoe (S):

$$MT = a + b \times \left(\frac{D}{W + S} + \frac{1}{2}\right)$$

This modification is due to the fact that Drury considered a target hit whenever any part of the participant's shoe touched it, effectively increasing the size of the target. Hoffman argued that ballistic and visually controlled foot movements should be modelled differently [116]. He proposed that while Fitts's law provides a good fit for visually controlled movements, ballistic movements are better modelled by the square root of the distance:

$$MT = a + b \times \sqrt{D}$$

Despite being originally created to model movement times, Fitts's Law can also be adapted to incorporate accuracy measurements. If end-point scatter data is available, this can be accomplished by using the effective distance (De) and effective width (We), where De is the mean movement distance from the start position to the end points and We is 4.133 times the standard deviation of the end points:

$$MT = a + b \times \log_2\left(\frac{D_e}{W_e} + 1\right)$$

Research so far has found little influence of foot dominance on movement times and accuracy. Chan et al. conducted a study in which they found no effect of gender of foot dominance on movement times in a reciprocal tapping task whilst seated [44].

We summarise pointing performance results for different input devices in Table 9.

Table 9 - Performance comparisons between hand the feet. Values correspond to ratios of task completion times and error rates for the feet versus the hand.

Ref.	Foot Device	Hand Device	Participants	$\frac{Foot\ Time}{Hand\ Time}$	$\frac{Foot\ Error}{Hand\ Error}$
[116]	None	None	10	1.95[1]	
[116]	None	None	10	1.7[2]	
[249]	Joystick	Mouse	17	2.32	1.56
[195]	Trackball	Trackball	9	1.6	1.2
[66]	Pedals	Tilt	24	1.05[3]	1.20[3]
[66]	Pedals	Touch	24	0.98[3]	1.87[3]
[89]	Joystick	Trackball	16	1.58[4]	

Notes: [1]: Ratio between the reported coefficients of the indices of difficulty for *visually controlled* movements.
[2]: Reported ratio for *ballistic* movements.
[3]: Ratio between reported means for selection time and error rate in the text formatting task.
[4]: Mean ratio between reported task completion times for the foot joystick and the mouse.

4.4.2.2 Reaction Times

Reaction time is the time elapsed between the presentation of a sensory stimulus and the corresponding behavioural response. A common model for reaction times is the Hick-Hyman law [109,124], which predicts that the reaction time for a set of n stimuli, associated with one-to-one responses is:

$$RT = a + b \times \log_2(n)$$

Simonen et al. compared reaction times of dominant hands and feet and found that reaction times were nearly the same in choice reaction time testing and the hand was slightly faster (125ms) than the foot in simple reaction time testing [243]. Reaction times are also dependent on the spatial mapping between foot controls and visual stimuli [45].

4.4.2.3 Learning Effects

An often overlooked issue in foot interaction studies is the learning effect. Since most users seldom use their feet for computing tasks, a significant amount of the performance gap between hands and feet could be explained by the lack of practice. In a study comparing the learning effects of a foot mouse and a hand trackball over ten sessions, Garcia and Vu found that while participants quickly reached a performance ceiling with the trackball, practice with the foot mouse significantly improved performance [88,89].

These results indicate that the feet suffer an unfair advantage in studies comparing hand- and foot-operated interfaces without allowing enough time for practice. In an early evaluation of pointing devices, in a time when users unfamiliar with a hand mouse still existed, a knee control fared comparably to the mouse, outperforming other hand interfaces [79].

4.4.3 Implicit

Regardless of how the feet are tracked the input can be used for explicit or implicit interaction. Schmidt distinguishes explicit from implicit interaction in that whereas in explicit interaction, "the user tells the computer in a certain level of abstraction (…) what she expects the computer to do", in implicit interaction, systems understand as input actions performed by the user that are not primarily aimed at interacting with a computerised system [232].

In Section 4.2.5 we showed that our lower limbs send nonverbal signals without our conscious knowledge. Analogously, several systems extract information from foot behaviour without us explicitly using this behaviour for interaction.

On an individual scale, several smart shoes implement this kind of interaction. By monitoring users from within the shoes, these systems are able to infer the user's identity [120], gait abnormalities [144] and monitor diseases such as diabetes [178].

On a public scale, augmented environments and surfaces are able to infer information about users unobtrusively by monitoring their feet posture. *GravitySpace* uses an FTIR-enabled floor to distinguish users, recognise poses and detect objects [31]. *Bootstrapper* recognises users by their shoes to personalise the usage of a multitouch table [222]. *Smart Floor* recognises users from their footstep profile [192].

Some works draw from Hall's theory of Proxemics [101] to make inferences about user's attention in regards to public displays. With laser range finders, Huber was able to distinguish users who were seeking information on a public display from those who were not solely based on their foot patterns [122]. Quek et al. also found high correlation between gaze orientation and feet position [215]. Considering that tracking the feet is arguably less invasive than the face or eyes, such approaches offer an interesting way to build context-aware public displays.

4.4.4 Multi-Modality

The feet serve one of two purposes in explicit interaction: as the main control of an application (primary) or as supporting other interfaces in the interaction (secondary). Typically, the feet are used as the primary modality for input when the user's hands are busy or dirty (e.g. [2]), the user has a disability or other accessibility issue that prevents him from using his hands (e.g. [249]) or it is awkward to reach the interface with the hands (e.g. [131]).

Most commonly, the feet are used to support the task being carried out by the hands. Previous research has explored combinations of the feet with several different modalities including a keyboard [89], a multi-touch table [230], gaze [91], tangible interfaces [10,233], large displays [62,235], a mouse [242] and a CAVE [154]. Typical secondary tasks assigned to the feet include acting as a modifier (e.g. guitar effects pedals, transcription pedals, foot switches as hotkeys), manipulation support (e.g. camera control in 3D environments [10] and mode selection (e.g. mode selection in a text editor [238] and in a musical keyboard [176]).

4.5 Discussion & Future Directions

As basis for our discussion, we use the categories previously assigned to the surveyed works as a framework for analysing the design space of foot-based interaction and provide directions for future work. Table 10 how the works described in the previous sections populate this design space.

Table 10 - Examples of instances of the design space of foot-based interactions

Pose	Sensing	Interaction	Examples
Sitting	Mediated	Semaphoric	[238]
		Deictic & Manipulative	[10,45,66,79,88,89,91,139,195,206,207,208,249,287,296]
	Intrinsic	Semaphoric	[196,262]
		Deictic & Manipulative	[287,290]
	Extrinsic	Semaphoric	[242]
		Deictic & Manipulative	[44,129,239,242,287]
		Implicit	[31,269]
Standing	Mediated	Semaphoric	[236]
		Deictic & Manipulative	[62,235]
	Intrinsic	Semaphoric	[2,56,154]
		Deictic & Manipulative	[50,110]
	Extrinsic	Semaphoric	[7,70,128,131,159,230]
		Deictic & Manipulative	[7,96,193,219]
		Implicit	[31,122,215,222]
Walking & Running	Mediated	Deictic & Manipulative	[64,126,264]
	Intrinsic	Semaphoric	[288]
		Deictic & Manipulative	[50,198]
		Implicit	[86,114,120,155,175,178]
	Extrinsic	Semaphoric	[104]
		Implicit	[31,119,192,281]

The first thing that is immediately noticeable in Table 10 is that the cell with the highest number of works is that of Sitting, Mediated, Deictic & Manipulative interaction. These works mostly study foot-operated computer peripherals for the desktop setting. Given that this is the most traditional HCI setting, the high popularity of this kind of work is understandable. However, few works look at long-term deployments of these interfaces, so more research needs to be done on the effects of practice on user pointing performance.

On the other hand, we also notice some empty cells (omitted from the table for space purposes).We found no works that investigate implicit interaction with mediated sensing, in any pose. Works that use mediated sensing to analyse human behaviour typically aim at understanding usage patterns or emotions, for example, by looking at the trajectory of the mouse [289] or how users press the buttons on a gamepad [259]. Therefore, we attribute the lack of studies of implicit user behaviour when interacting with foot-operated devices

to the limited number of use cases for such input devices and the small populations to which they are targeted.

Similarly, even though there are plenty of works that look at implicit interaction with intrinsic sensing when walking and running, we found no such works explicitly aimed at sitting and standing users. This is explained by the fact that the parameter usually being monitored by such smart shoes is the user's gait cycle. Possible directions for future research along these lines include using smart shoes to infer user states from their natural foot behaviour at their desks or even to infer conversation dynamics from interpersonal communication when standing upright.

From the table, we also notice that different poses lend themselves well to certain types of interaction. In the Sitting pose, Aside from Mediated, Deictic & Manipulative interactions, we also notice a large number of works investigating Extrinsic, Deictic & Manipulative interactions, evidencing a popularity of works investigating explicit control of desktop computers. In the Standing pose, we see a concentration of works using Extrinsic sensing. This shows that sensing the feet from the outside is well suited to applications in which the user interacts with a fixed installation, such as public displays or interactive floors. In the Walking & Running pose, we see a large number of works exploring Implicit interaction, evidencing that gait monitoring is a topic that has been explored in depth.

Based on the analysis of the papers reported in this survey, we achieved some insights, summarised in the following:

Feet excel at performing simple tasks. Feet lend themselves to performing simple tasks, such as operating a car's brakes. Yet, these tasks are as important as the tasks simultaneously performed by the user's hands. Several works in the literature mention that the feet are suitable for tasks where the precision of positioning is not of primary importance [195], but the feet are also capable of accomplishing highly complex tasks, such as playing the organ's baseline and manipulating three dimensional virtual objects [242]. However this requires a substantial amount of practice. The very few occurrences of such use cases suggest that hands—if they are not busy—are preferred for complex tasks.

Feet interfaces can assist the hands, rather than replace them. Traditionally, foot-based input was only successful in cases where hand-based input was not a viable option, either because the hands were already busy with other controls or because they were not available for other reasons (e.g. carrying objects, being dirty, etc.) [2]. This is arguably due to the feet being less dexterous, incapable of grasping objects the way our hands do, and being busy with locomotion. Therefore, it is not surprising that, for example, we are not using foot mice, but pedals are more widely adopted, such as in gaming or transcription. Nevertheless, a lot of research has been put into comparing the hands and the feet, leading to results where the hands consistently outperform the feet. Few works, however, explore how they can be used concurrently. We believe that the feet can effectively complement the hands, offering additional input channels with no homing time.

The performance of the feet might not be as bad as people think. Several works explored the possibility of reassigning cursor control from the hands to the feet, but in these experiments, the mouse consistently outperformed all foot interfaces (see Table 9). Garcia and Vu's results, however, suggest that this may not be because the feet are inherently bad for this purpose, but rather, because of lack of training [88,89]. Future work must take these learning effects into account, allowing users to become sufficiently familiar with the input device for a fair comparison with hand interfaces.

Foot-based input lends itself well to wearable computing. Wearable computing and augmented reality applications will benefit from foot-based input: users are on the go and interact spontaneously, with their hands often busy with other tasks (Section 4.1.2). At the same time, new sensing technology allows for capturing foot interaction with higher fidelity, allowing for both explicit and implicit interaction.

We still lack the understanding of how these interfaces work in the real world. Whereas there is a significant body of work on laboratory studies of foot-based interfaces, we still need to understand how these interfaces work when deployed for extended periods of time. Not only such in-the-wild studies could give us a better understanding of learning effects, it could provide insights on user acceptance of such interfaces.

Interactive systems can further benefit from what the feel tell about users' internal states. Even though the majority of the work in HCI that explores foot interfaces look into explicit interaction, by monitoring the feet, it is possible to recognise not only the activity in which the user engaged, but also internal states, such as attention, relaxation and anxiety (Sections 4.2.5 and 4.4.3).

4.6 Conclusion

In this survey we analysed foot interaction from the perspectives of the user, the systems and the interaction between users and systems. From the user perspective, we described the anatomy of the lower limbs and how it evolved. We then broke down the movements for each joint and related them to corresponding interaction techniques. We analysed the different poses in which users interact with foot interfaces — sitting, standing and walking/running — and how these poses influence the interaction. We then discussed how internal states are reflected by the behaviour of the lower limbs.

From the system perspective, we analysed how systems can use the feet to capture input and provide feedback to the user. We classified foot-operated sensing as mediated, intrinsic and extrinsic. We described actual implementations of such devices in the research literature and in commercial applications. We then discussed the different ways that such systems output information to the user.

Finally, from the interaction perspective, we categorised the foot actions into semaphoric, deictic, manipulative, and implicit. We compiled a dictionary of foot semaphoric gestures, aggregated foot performance results from previous work and described applications that infer internal states and activities from the posture and movement of the lower limbs. We then discussed applications that use the feet in combination with other input modalities.

Foot-based input substantially shaped the design of many devices and systems that matter to us and that we interact with on a daily basis, from cars to musical instruments. Whereas mechanical systems have employed foot controls for a long time, electronic devices are seldom designed to benefit from this modality. We believe that by better understanding the role of the feet in HCI, we can better design interfaces that take input from our whole body.

5 Empirical Investigations of Foot-Based Interaction

"Be sure you put your feet in the right place, then stand firm."

Abraham Lincoln

The literature review presented in the previous chapter highlights the fact that despite a lot of work having been done on foot-based interaction, there is still a lack of understanding of the most fundamental aspects of the interaction process. In this chapter, we take an empirical approach to investigate such basic aspects of foot movement for desktop human-computer interaction. First, we describe a custom-built foot tracker based on a Kinect sensor mounted under the desk developed to enable our investigation. We then describe the results of four user studies that examine different aspects of the interaction. In the first one, we use this tracker to derive one-dimensional and two-dimensional Fitts's Law models of foot pointing performance. In the second, we investigate whether users' pointing performance is affected by the direction of movement. In the third, we investigate how the feet can be used for simultaneous manipulation of two parameters. In the fourth, we explore the parallel use of hands and feet for manipulating multiple parameters.

Computer interfaces operated by the feet have existed since the inception of HCI [79], but such devices remained restricted to specific domains such as accessible input and audio transcription, being largely overshadowed by hand-based input in other areas. However, this overshadowing cannot be put down to lack of dexterity, as we regularly accomplish a wide variety of everyday tasks with our feet. Examples include the pedals in a car, musicians' guitar effect switches, and typists' use of transcription pedals. Recent technological advances renewed interest in foot-based input, be it for interacting with a touch-enabled floor [7], for hands-free operation of mobile devices [236], or for adding more input channels to complex tasks [242]. Despite this, we still lack a thorough understanding of the feet's capabilities for interacting in one of the most common computing setups—under the desk.

In particular, unlike previous work that used trackballs [195], pedals [66], and foot mice [88], we wished to explore unconstrained feet movements. This removes the need for a physical device (as well as the related foot-to-device acquisition time) and provides a wide range of interaction possibilities (analogous to the ones available from a touch-screen over a mouse).

We envision numerous applications to arise from this greater understanding. These include using your feet to scroll a page while the hands are busy with editing the document, changing the colour of a brush while moving it with the mouse, or manipulating several audio parameters simultaneously (using both the hands and the feet) to create novel musical performances.

To address this gap we conducted a series of experiments exploring different aspects of foot-based interaction. In the first, we recorded 16 participants performing 1D and 2D pointing tasks with both feet to build the first ever ISO 9241-9 Fitts's Law models of unconstrained foot pointing for cursor control. This first study provided some evidence that side-to-side movement is faster than backwards and forwards. To confirm this hypothesis, we conducted a second experiment in which participants performed 1D serial pointing tasks in each direction. In the third experiment, we investigated the manipulation of multiple parameters using one and two feet. In the fourth and final experiment, we evaluated the use of the feet together with the hand.

In summary, (1) we built 1D and 2D ISO 9241-9 compliant movement time models for unconstrained foot pointing; (2) we found that unconstrained foot pointing is considerably slower than mouse pointing, but comparable to other input devices such as joysticks and touchpads; (3) we found no significant difference in performance between the dominant and non-dominant foot; (4) we found that left and right movement is easier than backwards and forwards; (5) the most comfortable movement for desktop foot interaction is heel rotation; (6) techniques that have a direct spatial mapping to the representation outperform the others; (7) when variables are shown separately, two feet work better than one; (8) we show that the feet perform similarly to the scroll wheel in tasks where the feet are used in conjunction with the mouse; and (9) we provide design guidelines and considerations based on our findings.

5.1 Tracking the Feet under the Desk

To track users' feet we implemented a prototype based on the depth data from a Microsoft Kinect mounted under a desk (see Figure 28). Our system tracks the three-dimensional positions of the foot and ankle joints in relation to the floor, as well as the orientation of the foot. Richter et al. also use a Kinect to detect the users' feet; in their work, the authors were concerned in isolating the feet to identify users standing by a multitouch table, rather than tracking for interactive purposes [222]. Hu et al. also used a custom-built foot tracker based on a Kinect, but their algorithm, despite providing a highly accurate model, does not work in real time. System calibration is not user dependent, so it only needs to be calibrated when the setup changes. In this step the system takes a snapshot of the background without the user's feet. The user then draws on the image of the background a rectangle that will define the active area of the tracker. We use three corners of this rectangle (lower left, upper left, and lower right) to extract the floor plane and define the new basis for the coordinate system. The axes of the new coordinate system are formed by the normalised vectors parallel to the two edges formed by the lower left vertex with its adjacent vertices, and their cross-product. This vertex becomes the new origin. We then convert the data of subsequent frames to this new coordinate system. This change of basis allows more freedom in the camera positioning, as the system knows the position and orientation of the camera relative to the floor, allowing the camera to be positioned in a more suitable height or orientation depending on the physical configuration of the desk.

Figure 28- Experimental setup for my studies. Participants sat at the desk as they normally would, while their feet was tracked by a Kinect under the desk.

To interact with the system, the user places the feet in the tracked area (see Figure 29a), and for every depth frame captured by the system (see Figure 29b), we convert the coordinate system (see Figure 29c) and subtract the background frame, hence isolating the user's legs and feet (see Figure 29d). We then isolate the feet from the legs by thresholding the depth data at 0.1m above the floor (see Figure 29e). We finally fit ellipses that have the same second moment to the thresholded image of each foot and use their foci as the positions for the feet and ankles (see Figure 29f).

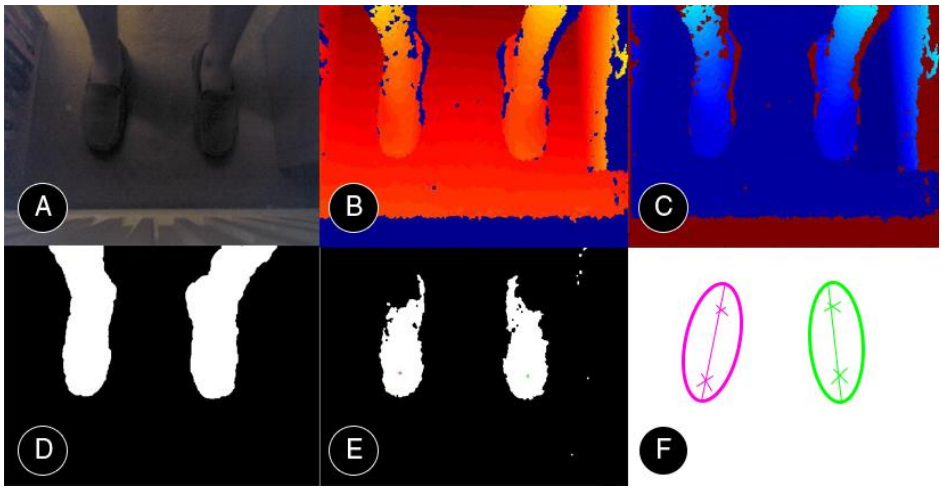


Figure 29 - Feet tracking algorithm: (a) Colour image; (b) Raw depth, relative to the camera; (c) Depth, relative to the floor; (d) Legs isolated from the background; (e) Feet isolated form the legs; (f) Ellipses fitted to the mask. The foci of the ellipses are used as the joint positions for the feet and ankle.

A problem commonly encountered in vision-based systems is lighting conditions and the area under the desk is particularly problematic in this sense as it is often very poorly lit. Because our approach does not rely on the colour feed from the camera, it works well in different lighting conditions. Another advantage of our approach is that we do not take previous tracking states into account, so if the system loses track of the feet momentarily

(e.g. if the user leaves his desk), it can instantly recover from erroneous states. Also, it does not need to be calibrated for different users. We implemented our tracker in Matlab, and it runs at a frame rate of approximately 25 frames per second on a Windows 8.1 PC with 32GB of RAM, with an Intel Core i7 CPU @2.90GHz.

5.2 Models of Foot Pointing Performance

In Chapter 4, we summarised previous works that attempted to quantify the performance of hand and foot pointing. These works evaluated several different foot interfaces, but the wide variety of experimental designs makes comparing results difficult. Further, these studies have only looked at 1st order devices (i.e. devices that control the rate of change of a value, rather than the value directly, such as the joystick and the pedals) and relative input devices (i.e. devices that sense changes in position, such as the mouse and the trackball) [111]. Hoffman investigated unconstrained absolute positioning, but with users tapping on physical targets rather than using the foot for cursor control. However, modern devices that take input from the feet, such as depth cameras [116] and interactive floors [7]use absolute positioning, making it important to study this kind of interaction.

To fill this gap, we conducted an experiment in which 16 participants performed 1D and 2D pointing tasks with both feet, to build the first ever ISO 9241-9 Fitts's Law models of unconstrained foot pointing for cursor control. This allows us to compare our model to those of other input devices based on the same standard using the mean throughput for each condition. We also tested for effects of task and foot on user performance.

5.2.1 Participants

We recruited 16 participants (11M/5F), aged between 20 and 37 years (median = 27), with foot sizes ranging from 23 to 30cm (median = 26cm). Participants were inexperienced with foot interfaces and half of them were regular drivers. All participants were right handed, but one was ambidextrous. All participants were right footed.

STUDY AT A GLANCE

Goal: Build performance models for unconstrained feet movements and test for the effects of foot and task

Method: Within-subjects experiment

Participants: 11M/5F (20-37y.)

Independent variables: Foot side and task (1D or 2D)

Dependent variable: Throughput

Results: No difference between feet, 1-D is faster than 2-D

5.2.2 Apparatus

The experiment was conducted in a quiet laboratory space, on a laptop with an 18-inch screen and 1920×1080 resolution. To track the feet, we used the foot tracker described in section 5.1. We made sure that only one foot was visible to the camera at any point in time, by asking participants to keep the opposite foot under the chair, and that the cursor control was assigned to the toes of whichever foot was in view. Mouse clicks were performed using a conventional mouse with disabled movement tracking. The tracker was calibrated with a 1:1 CD gain, so that the cursor and foot movements matched exactly.

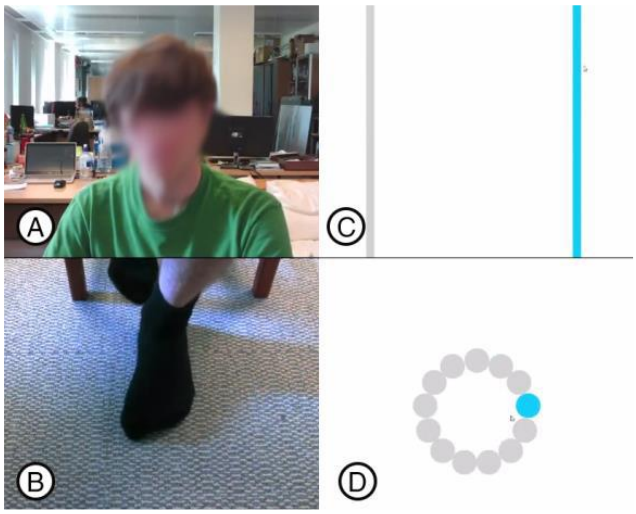


Figure 30 - We recorded participants faces (A), and feet (B), synchronised with the 1D (C) and 2D (D) Fitts's Law tasks.

Participants performed 1D and 2D Fitts's Law tasks, for which we used Wobbrock et al.'s *FittsStudy* tool, an ISO-9241-9-compliant C# application to "administer conditions, log data, parse files, visualize trials and calculate results" [286]. The tool was configured to administer nine different combinations of A (amplitude) × W (width) defined by three levels of A {250, 500, 1000} crossed with three levels of W {20, 60,130}, yielding nine values of ID {1.55, 2.28, 2.37, 3.12, 3.22, 3.75, 4.14, 4.7, 5.67}.

We recorded all sessions using additional cameras pointed at participants' faces and feet, as well as the screen using the *Open Broadcaster Software* (see Figure 30).

5.2.3 Procedure

Participants first signed a consent form and completed a personal details questionnaire. The tasks were conventional ISO 9241-9 pointing tasks, in which targets appeared in blue on the screen. Participants selected targets by moving their feet so the cursor was above the target and by left-clicking the hand-held mouse. We chose this technique rather than foot tapping as we were interested in the time it takes to move the feet and the gesture time might delay the task unnecessarily.

Participants performed both a 1D task (with vertical ribbons on either side of the screen) and a 2D task (with circular targets in a circular arrangement), with both their dominant and non-dominant foot. The order of tasks was randomized, but we ensured that the same foot was not used twice in a row. Each task comprised 9 IDs and was repeated in 13 trials (the first 3 discarded as practice). To summarize, each participant performed 2 feet × 2 tasks × 9 IDs × 13 trials = 468 movements. To make the friction with the floor uniform across all users, we asked them to remove their shoes and perform the tasks in their socks.

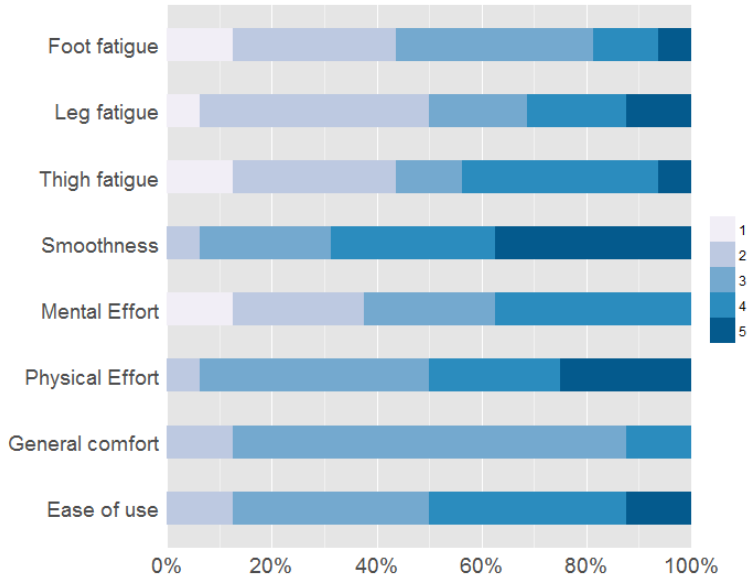


Figure 31 - Distribution of responses for the subjective reactions to the interaction technique (1-Low, 5-High). We modified the ISO 9241-9 questionnaire for the feet.

After completing the tasks we asked participants to fill in a questionnaire adapted from the ISO 9241-9 standard for the use with the feet (see Figure 31). We also conducted an open-ended interview about participants' experience using the foot interface, what they liked and disliked about it, what strategies and movements they used to reach targets, etc. All interviews were transcribed and coded accordingly.

5.2.4 Results

Our analysis had two objectives: to build a Fitts's Law performance model for each task and each foot and to check whether there was any difference in performance—as measured by the throughput—for different feet and tasks.

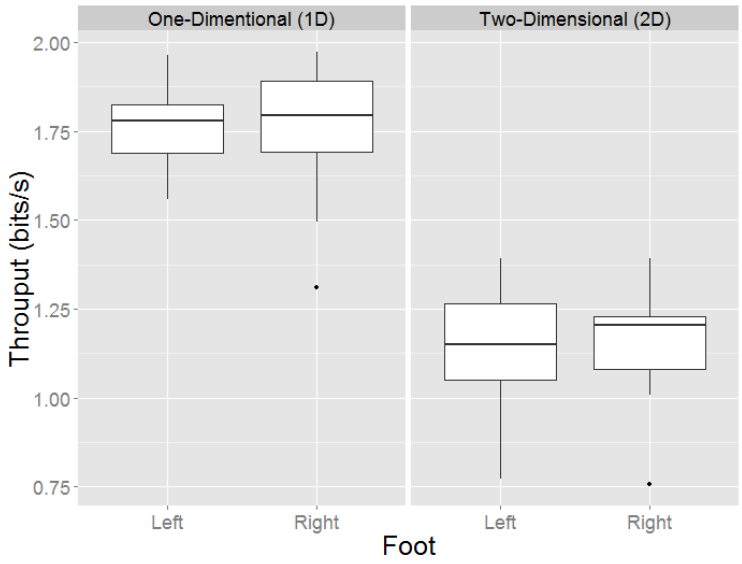


Figure 32 - Distribution of mean throughputs per participant for each condition

To build the performance models, we computed the mean movement time (MT) and the mean 1D and 2D effective indices of difficulty (IDe) for each participant and for each combination of A×W. We then built the performance models using linear regression on these data, using the formulation described by Soukoreff and MacKenzie [247]. Table 11 summarizes the movement time models, as well as the R-squared and the mean throughput averaged over the individual mean throughputs for each participant. To test for differences in performance for each condition, we compared the mean throughput (see Figure 32)—a metric that takes into account both speed and accuracy of the movement performance [247]—using a factorial repeated-measures ANOVA. We found a significant main effect of the task on the average throughput ($F(1,15) = 391, p < .001$), with an average TP of 1.75 for the 1D task and 1.15 for the 2D task, but not of the foot ($F(1,15) = .75, p = .09$) or the interaction between foot and task ($F(1,15) = .41, p = .12$).

Table 11 – Performance model for each condition with its corresponding r-squared, mean throughput and error rate

Condition	Movement Time (ms)	R^2	Throughput (bit/s)	Error Rate (%)
1D Right	$99 + 561 \times ID$	0.88	1.75	8.43
1D Left	$-56 + 609 \times ID$	0.96	1.75	8.49
2D Right	$423 + 739 \times ID$	0.85	1.16	7.64
2D Left	$372 + 789 \times ID$	0.64	1.14	8.60

After transcribing and coding the interviews, some consistent patterns of users' opinions emerged.

Movement Behaviours: In general, participants preferred to move around their hips and knees as little as possible, leaving as much of the movement as possible to the ankle joints. Participants reported five strategies for reaching targets on the screen: dragging the foot, lifting the foot, rotating the foot around the heel, rotating the foot around the toes and nudging the toes. At the beginning of the tasks, participants often started by dragging the foot across the floor, but quickly realized that this was tiring ("*was a bit uncomfortable*", "*I could instantly feel my abs working*", "*more taxing and not really natural*"). Four participants reported lifting the foot across the floor, but found that keeping the foot up was rather tiring ("*I'd have more control and I don't have the friction of the surface, but then I got very fatigued from keeping my whole leg up*").

These strategies were used when targets were far apart; for shorter distances participants reached the targets by rotating the foot around the heels with the toes up, what they often referred to as "pivoting" ("*most of the time, I just tried to move around my heel*"). The reported advantages of heel rotation were the ease of movement, less fatigue, higher comfort and higher precision. Finally, for small adjustments and smaller targets, participants employed the toes in two ways: one participant reported rotating the foot around the toe and six participants reported bending and extending their toes, which would nudge the cursor towards the target ("*when I wanted to do a fine grained, on the smaller targets, I would crunch my toes*").

Differences between Tasks: All participants but one found the one-dimensional task easier than the two-dimensional one, which is reflected in the quantitative difference in throughput. This can be explained by the fact that moving left and right could be accomplished with heel rotation (the easiest movement, as participants reported), whereas back and forth movements required knee flexion and extension, either by dragging the foot on the floor or lifting it above it, both strategies that were reported as being tiring.

Challenges: The biggest challenges reported by participants were the cognitive difficulty in reaching small targets ("*when the targets are smaller you need more precision so you need to focus*") and in coordinating the hands and feet ("*it was weird starting, because you'd have to coordinate your thought process, your clicking and your feet, but I think as you went on, It was pretty quick to adapt*"), fatigue ("*a little fatigue influenced the outcome*"), friction with the floor ("*I don't like this kind of rubbing with the floor*"), and overshooting ("*I knew that I was going to overshoot, so I just overshot and tried to click at the same time*").

5.2.5 Discussion

Our regression models are in line with previous work with our one-dimensional model being very similar to Drury's ($MT = 189 + 550 \times ID$) [71]. Hoffman found a much lower coefficient ($MT = -71 + 178 \times ID$) [116], but both him and Drury conducted experiments with physical targets rather than cursor control. As Drury noted, this effectively increases the sizes of the targets by the size of the participant's shoe [71]. Also, whereas we use the Shannon formulation of ID, Drury used Fitts's original formulation and Hoffman used the Welford formulation.

Since our model is compliant with the ISO standard, we can compare our throughputs with other studies reported in literature. The typical range of throughput for the mouse is between 3.7 and 4.9 bit/s, considerably higher than the 1.2–1.7 range we found for the feet, but expected given users' experience and practice with it [247]. The values we found, however, fall into the range for other input devices such as the isometric joystick (1.6–2.55) [247], the touchpad (0.99–2.9) [247] and video game controllers (1.48–2.69) [182].

By allowing participants to choose how to reach the targets, we obtained valuable insights into the most comfortable ways of using the feet. Although heel rotation was perceived as the most comfortable movement, most foot-operated interfaces do not use this movement (an exception is Zhong et al.'s *Foot Menu* [296]). Our results are also in line with Scott et al.'s in which users also reported that heel rotation was the most comfortable gesture, followed by plantar flexion, toe rotation and dorsiflexion [237]. The use of heel rotation is suitable for radial and horizontal distributions of targets. This kind of interaction could be used in a discrete (e.g. for foot activated contextual menus) or in a continuous fashion (e.g. controlling continuous parameters of an object while the hands perform additional manipulations).

We investigated foot performance for seated users so our results apply to foot-only (e.g. mice for people with hand disabilities), and foot-assisted (e.g. driving simulators, highly-dimensional applications) desktop interfaces. It remains to be seen how these results apply to standing users (e.g. using a touch-enabled floor). A second limitation is that our participants were not familiar with this kind of input device, which might affect the predictive power of our models if the device is used more frequently.

5.2.6 Conclusion

We presented ISO 9241-9 performance models for 1D and 2D foot pointing in a sitting position. Our results suggest little difference in performance between the dominant and non-dominant foot and that horizontal foot movements are easier to perform than vertical ones. We identified five strategies that participants used to reach targets and found that the preferred one was rotating the foot around the heel. We also found that the biggest challenges for foot-based interaction in a desktop setting are difficulties in reaching small targets, hand-feet coordination, fatigue, friction with the floor, and overshooting targets. These findings are important because they help us complete our understanding of the potential of foot-operated interfaces and provide guidance for future research in this emerging domain.

5.3 The Effect of Direction on Foot Pointing Performance

One possible use for the feet in a seated position is to provide one-dimensional input, be it discrete (e.g. selecting an option in a menu) or continuous (e.g. changing the music volume). To better understand how to better design such interfaces, it is important to understand if there is a significant difference in the movement times and comfort across different directions. Therefore, we conducted a serial one-dimensional Fitts's Law experiment. In this study, we tested the effects of the direction in which the targets were distributed (horizontal vs. vertical) and the foot (dominant vs. non-dominant) on the movement times and error rates.

5.3.1 Participants

For this experiment, we recruited 10 participants (8M/2F), aged between 19 and 31 years (mean 27), with posters on campus and adverts on social networks. All participants were right-handed and one was left-footed; seven participants were car drivers. Foot sizes ranged from 22 cm to 33 cm (mean 27.1 cm). None of the participants had ever used a foot mouse or similar foot-operated pointer before.

STUDY AT A GLANCE

Goal: Compare the performance of foot pointing in the x and y axes.

Method: Within-subjects experiment

Participants: 12(8M/4F)

Independent Variables: Direction and Foot Dominance

Dependent Variables: Throughput

Results: No effect of foot dominance, but movement in the x axis is faster than in the y axis.

5.3.2 Apparatus

The experimental setup was the same as in the previous study (see Figure 28). Participants sat facing the monitor on a fixed chair. To make sure that the friction with the floor was the same across all participants, as well as to remove the additional friction from the shoe sole, we asked participants to take off their shoes and wear socks during the experiment. The system was calibrated to track an area of 50 cm × 50 cm under the desk.

To begin, participants signed a consent form and filled in a questionnaire. The task in our experiment was analogous to other Fitts's Law experiments. The user was presented with a green and a red bar with a certain width (W) and separated by a certain distance (A) on the screen. For each trial, the user had to select the green bar, at which point the colours of the bars switched.

To select a target, participants used their feet to position an on-screen cursor over the green bar and press the space bar. We chose to perform the activation this way rather than with a foot gesture, such as a tap or a touch on the floor, for several reasons: first, the movement responsible for the activation might interfere with the positioning of the cursor. Second, the gesture could influence the movement of the feet.

For example, in order to select a target the user had to touch the corresponding point on the floor, he would be forced to move the cursor by lifting his foot off the floor, so as to minimise accidental activation. Conversely, if the selection was performed by tapping, dragging the foot on the floor would seem more intuitive. By using the hands to perform the activation, we freed participants to use whichever strategy they felt more suitable or comfortable for the task. We selected 14 combinations of W and A to yield exact indices of difficulty from 1 to 7, using the Shannon formulation. Each ID combination was executed with each foot in both horizontal and vertical configurations. We balanced the order of the feet and the direction of the bars among participants but and ensured the task was not repeated with the same foot twice in a row so as to reduce fatigue. The order of difficulty was randomised.

The complete procedure was repeated ten times. To summarise, each participant performed 2 feet ×2 directions×14 ID combinations×10 repetitions = 560 movements.

The system continually logged the position of the feet and cursor and a video camera placed under the desk recorded participants' leg movement. At the end of the experiment participants filled in another questionnaire on the perceived difficulty and speed of the target selection on the top, right, bottom, left and centre of the screen for each foot. We also asked for suggested applications of foot-operated interfaces.

5.3.3 Results

To ease reporting, we use the following abbreviations: horizontal movement with the right foot (RH), vertical movement with the right foot (RV), horizontal movement with the left foot (LH) and vertical movement with the left foot (LV).

Table 12 – Performance model for each combination of foot and direction as well as the r-squared and mean throughput.

Condition	Movement Time (ms)	R^2	Throughput (bit/s)
Right Horizontal	$128 + 449 \times ID$	0.66	2.20
Left Horizontal	$102 + 497 \times ID$	0.66	2.02
Right Vertical	$-140 + 606 \times ID$	0.55	1.94
Left Vertical	$-138 + 608 \times ID$	0.52	1.94

Following the recommendations provided by Soukoreff and MacKenzie [247] we adjusted the data for accuracy, computing the Effective Index of Difficulty for each participant, for each condition, obtaining 140 pairs of IDe and MT for each group. From this data, we obtained the four Fitts's Law regression models for our experiment (see Table 12), using the Shannon Formulation and the effective indices of difficulty. The logarithmic value of the index of difficulty significantly predicted the performance time for all conditions. The overall model with the logarithmic value of the index of difficulty also predicted the performance time well for RH ($R^2 = 0.66$, $F_{1,138} = 268.4, p < .001$), LH ($R^2 = 0.66, F_{1,138} = 269.0, p < .001$), RV ($R^2 = 0.55, F_{1,138} = 167.9, p < .001$) and LV ($R^2 = 0.52, F_{1,138} = 149.2, p < .001$).

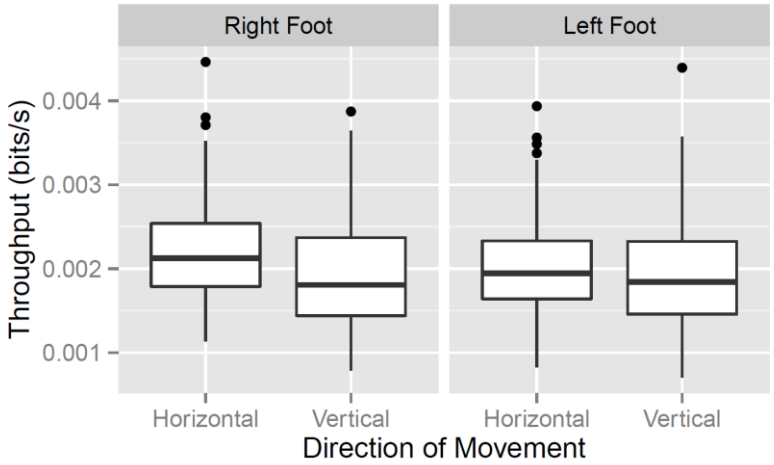


Figure 33 - Throughput for each foot in the horizontal and vertical tasks

To compare the models we computed the mean of means of the throughput of each condition. The throughput is a measure that encompasses both the speed and accuracy of the movement. The values for the mean throughput in each condition are shown in Table 12. We compared the throughputs using a factorial repeated-measures ANOVA. We found a significant main effect of the direction of movement, $F_{1,11} = 14.06;\ p < .05$, but not of the dominance of the foot used, $F_{1,11} = 4.62, p\ = .081$, on the task completion time. We also found no significant interaction effect between the foot and the direction of movement, $F_{1,9} = 4.72, p = .052$ indicating that both feet perform roughly the same in both directions. Figure 33 shows the throughputs for each condition. These results suggest that it does not matter which foot is used, but moving it horizontally is faster than moving it vertically.

Participants were consistent in their strategy for reaching targets: they would position their foot in the general target area and perform fine positioning in a second step. During vertical movement where the foot moved forward and backward, participants nudged their foot, wiggled their toes and flexed the foot for fine-positioning. For horizontal movement, users consistently reported pivoting the foot around the heel, only dragging if the target could not be reached by pivoting.

Most users (70%) found using their dominant foot to select the bars on the left and right of the screen the easiest and most (also 70%) found using the non-dominant foot to select targets on the upper and lower edges the hardest task. We asked users for the hardest aspect of using the interface and responses indicated that fine-grained positioning is the biggest problem.

Suggestions for tasks that could be improved by the use of the feet together with traditional input modalities pointed to the fact that it is not suitable for fine positioning, but it would be useful for **mode switching** *("switching tasks", "switching between colours when drawing", "changing tabs in a browser"*), **navigation** *("scrolling", "game exploration", "Google maps", "navigating a document"*), and **selection between a reduced number of options** *("anything where you have a limited number actions to do", "two or three big buttons", "if there were large quadrants, it would be useful"*). For real-world use, one participant said that he "(...) *would not want the tracking to be always on. To toggle this mode, I would suggest holding down a key*".

5.3.4 Discussion

The intercepts and slopes shown in Table 12 are in line with previous Fitts's Law studies. Soukoroff and MacKenzie suggested that intercepts should be between -200 ms to 400 ms and ours lied within this range. Further, none of the intercepts were significantly different than zero (p > .05), which is expected for a theoretical ID of zero. We therefore refrain from making any assumptions about the meaning of our intercepts and attribute the values to subject variability. We can also compare our models to those of other Fitts's Law studies using a mouse in point-select tasks. In MacKenzie's comparison of six such models [165], the slopes varied from 83 to 430 bits/ms, with bandwidths ranging from 2.3 to 12.0 bits/s. This makes the fastest group in our study (horizontal movements with the dominant foot) between 5% and 80% slower than the hand. This is also in line with the findings of Pakkanen and Raisamo, who found the feet to be 60% slower than the hands [195].

Our results indicate that when manipulating one parameter with the feet, horizontal movements are faster than vertical movements and users tended to pivot their feet rather than drag them. This suggests that when designing controls for the feet, radial layouts might be more efficient than linear layouts.

5.4 Simultaneous Manipulation of Two Parameters

The previous experiment focused on how the feet can be used to control one parameter. However, the feet have a greater bandwidth than one parameter as their positions and orientations in space can have meaning for input. In this experiment, we aimed to understand how people can use their feet to control two parameters at the same time. Is it better to use one foot to control multiple parameters or distribute these parameters across the two feet? Further, does the visual representation of the control of parameters affect the interaction?

STUDY AT A GLANCE

Goal: Compare different foot-based interaction techniques for controlling multiple parameters

Method: Within-subjects experiment

Independent Variables: Interaction technique and visualisation

Dependent Variables: Completion time and error rate

Results: Fastest technique was using one foot in the rectangle visualisation.

5.4.1 Participants and Apparatus

For this experiment, we recruited a group of 12 participants (8M/4F), aged between 19 and 42 years (mean 28) using posters on campus and adverts on social networks. Two of the participants were left handed and footed and nine were drivers. Foot sizes ranged from 22 cm to 32 cm (mean 26.7 cm). None of the participants had ever used a foot mouse or similar foot-operated pointer. The experimental setup was the same as for the previous experiment.

5.4.2 Procedure

Participants were first asked to sign a consent form and complete a personal information questionnaire. They were then given time to familiarise themselves with the interface. The goal of the study was to investigate how interaction technique and visualisation influence task completion time and error rate. To this end, participants were asked to manipulate two variables, within a certain threshold, while we varied the following two factors:

- INTERACTION TECHNIQUE (3 levels): The two input values were manipulated by (1) XY position of 1 foot (1F); (2) X position of both feet (XX) and; (3) X position of one foot and Y position of the other (XY).
- VISUALISATION (2 levels): Rectangle resizing and slider adjustment (described below).

In the first visualisation, the task was to fit the dimensions of an adjustable rectangle to those of a target rectangle (see Figure 34a). The target values were the width and height of the destination rectangle while the threshold was represented by the thickness of the rectangle's stroke. In the second visualisation, participants were asked to set two sliders along a scale to different target values marked by red tags (see Figure 34b). Here, the target values were the centres of the tags and the threshold was represented by their thickness. We chose these two visualisations because in the first the two values were integrated (as the corner of the rectangle) while in the second the values were represented independently (as separate sliders). We hypothesised that these different visualisations might influence the performance depending on the number of feet used in the interaction. For each task, we measured the task completion time and error rate.

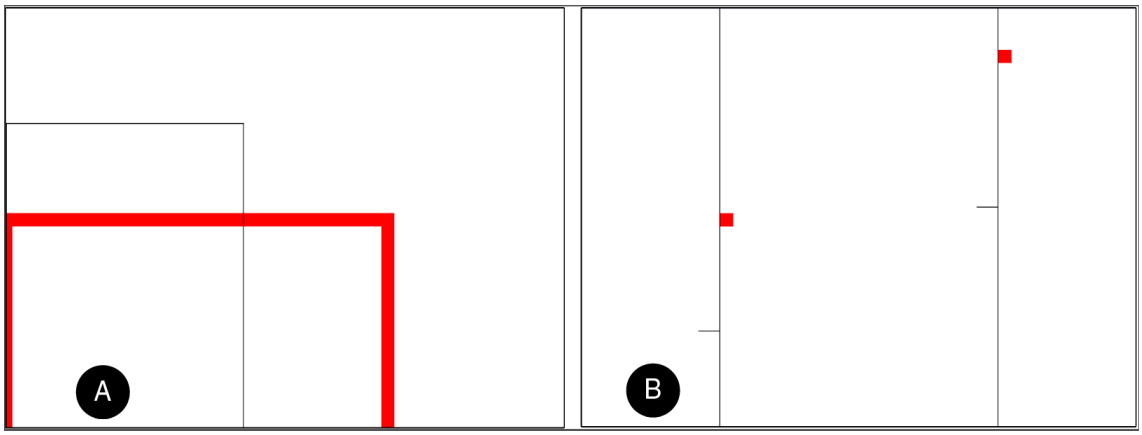


Figure 34 - Tasks in the third experiment, using the rectangle resizing (A) and slider matching (B) visualisations.

To summarise, each participant performed: 3 interaction techniques × 2 visualisations × 30 repetitions (varying the target values and thresholds) = 180 tasks.

5.4.3 Results

Table 13 shows the mean task completion time and error rates for each condition. We compared the task completion times (see Figure 35) using a factorial repeated-measures ANOVA. Mauchly's test indicated that the assumption of sphericity had not been violated neither by the effects of the technique ($W = 0.88, p = 0.64$) nor by the effects of the interaction between technique and visual representation ($W = 0.53, p = 0.11$). Sphericity was not an issue for the effects of the visual representation because it had only two levels. All effects were reported as significant at $p < .05$. There was a significant main effect of the technique, $F_{2,22} = 14.82$ and of the visual representation, $F_{1,11} = 50.46$ on the task completion time. There was also a significant interaction effect between the technique and the visual representation, $F_{2,22} = 34.10$ indicating that the interaction technique influence on participants' speed was different for the rectangle and slider representations of the task.

Table 13 - Mean task completion time and mean error rates for each condition in experiment 3.

	Rectangle			**Sliders**		
	1F	XX	XY	1F	XX	XY
Time (ms)	3428	5063	4161	9098	4625	4902
Error Rate (%)	5.5	3.7	8.0	8.2	8.1	8.0

Bonferroni post-hoc tests revealed that using one foot is significantly different than all other conditions in the slider representation ($p < .05$), but not in the rectangle one, as in this condition, it was not significantly different than using one foot horizontally and the other foot vertically ($p = 0.38$). The two conditions in which participants used both feet were not significantly different in any combination of techniques and representations at $p < .05$.

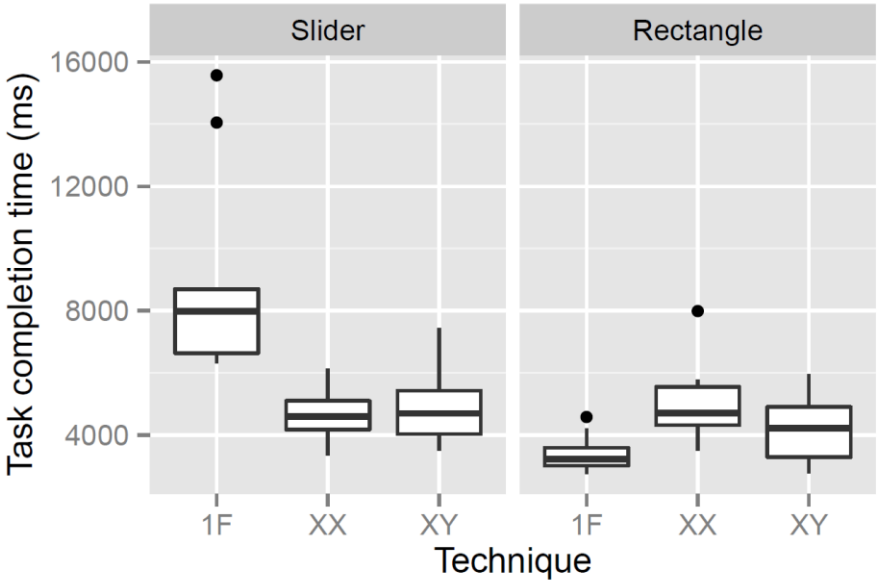


Figure 35 - Movement times for each task (resizing the rectangle and setting the sliders) and technique (one foot - 1F, two feet horizontally - XX, one foot horizontally and one vertically - XY)

We also compared accuracy using a factorial repeated-measures ANOVA. Mauchly's test indicated that the assumption of sphericity had not been violated neither by the effects of the technique ($W = 0.78, p = 0.42$) nor by the effects of the interaction between technique and visual representation ($W = .91, p = .72$). Our results showed no significant effect of the technique ($F_{2,16} = 0.21$, $p = 0.82$), of the visual representation ($F_{1,8} = 4.47$, $p = 0.067$) or the interaction between them ($F_{2,16} = 0.14$, $p = 0.87$).

5.4.4 Discussion

Our results show that when manipulating multiple variables with the feet the visualisation strongly affects performance. The best performances amongst all conditions were interaction techniques 1F and XY in the rectangle representation, which were not significantly different at $p < .05$. In these two conditions, there was a direct spatial mapping between the technique and the task, since in technique 1F, the foot moved together with the corner of the rectangle and in technique XY, the feet moved together with its edges.

Users were confused when this spatial mapping was broken. The worst performing condition was using technique 1F for the slider task. Even though the underlying task was exactly the same, the change in visualisation caused the mean completion time to increase over twofold. This can be explained by how users would complete the task. In the slider task, participants would often set one slider at a time and in technique 1F, this meant moving the foot in one direction and then in the other. The problem is that users find it hard to move the foot in only one direction at a time. As we discovered in our previous study, when moving the foot horizontally, users tend to pivot their feet, rather than drag them, and this movement causes the cursor to move in both directions at the same time, resulting in users setting one slider, then setting the second one and having to go back and forth between them to make final adjustments. This was not a problem when controlling each value by a different foot. Regardless of whether the user tried to set both values at the same time or in sequence, moving one foot did not affect the other, so the visual representation was not an issue when using two feet.

An interesting effect we observed was in technique XY in the rectangle representation. Even though only one axis of the movement of each foot was being used to control the size of the rectangle, some participants would move both feet diagonally and symmetrically. One participant was even conscious of this, but kept on using this strategy: *"I knew that each foot controlled only one dimension, but I found myself moving each one in both directions."* This suggests that symmetrical movements might be more comfortable than independent ones when using two feet.

5.5 Parallel Use of Feet and Hands

The previous experiments investigated interactions using the feet alone. In this experiment we wanted to investigate the overhead caused by using the feet in parallel with one hand. More specifically, we wanted to test whether there is an effect of resizing technique (scrolling with the mouse wheel, the position of one foot, or the distance between two feet) on the completion times and accuracy of the task, while the hand repositions the same square.

5.5.1 Participants and Apparatus

For this experiment, we recruited a group of 13 participants (11M/2F), aged between 19 and 32 years (mean 26), with posters on campus and adverts on social networks. Two participants were right left and footed while 10 participants were drivers. Foot sizes ranged from 22 cm to 34 cm (mean 28 cm). None of the participants had ever used a foot mouse or similar foot-operated pointer. The experimental setup was exactly the same as the one for the previous experiments.

STUDY AT A GLANCE

Goal: Investigate the overhead of using the feet in parallel with the hand.

Method: Within-subjects experiment

Participants: 12 (10M/2F)

Independent Variables: Interaction technique

Dependent Variables: Completion time and error rate

Results: No significant difference between the techniques in time, but the scroll wheel yielded fewer mistakes

5.5.2 Procedure

Upon arrival, participants signed a consent form and completed a personal information questionnaire. They were then given some time to familiarise themselves with the interface. The task consisted of resizing and positioning a square to match a destination square at a different place on the screen. In all experimental conditions, the positioning was done with the mouse but the size of the square would be manipulated by one of three controls: the scroll wheel of the mouse, the horizontal coordinate of one foot or the horizontal distance between the two feet. We chose the scroll wheel as it is widely used for manipulating continuous variables. When the size and position of the two squares were matched, the user would click with the mouse and the button would reappear in the centre of the screen. Each participant repeated this task 40 times for each condition, with the target square in different positions and with different sizes. We measured the task completion time and the error rate. In the end of the study participants were asked to rank their preference of interaction techniques.

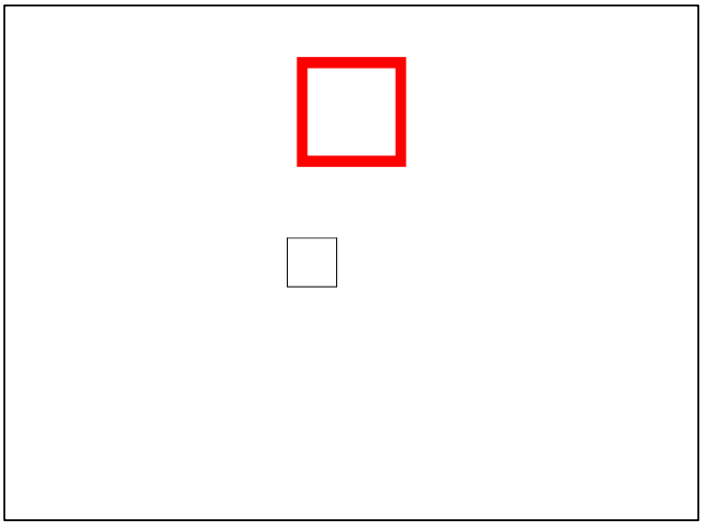


Figure 36 - Task in the fourth experiment. Participants were asked to translate and resize the black rectangle to match the red one.

5.5.3 Results

Figure 37 shows the mean task completion times and error rates for each condition. The mean task time was similar across all conditions: 2.50s for the scroll wheel, 2.63s for the two feet condition, and 2.95s for the one foot condition. We first compared the task completion times in each condition with a one-way repeated measures ANOVA. Mauchly's test indicated that the assumption of sphericity had not been violated, $W = 0.55, p = 0.05$. Our results showed a significant effect of the technique used for resizing on the task completion time, $F_{2,22} = 5.08, p < .05$. Post-hoc tests showed that using one foot was significantly slower than the other two conditions ($p < .05$), but no significant difference was found between using two feet and scrolling ($p = .062$).

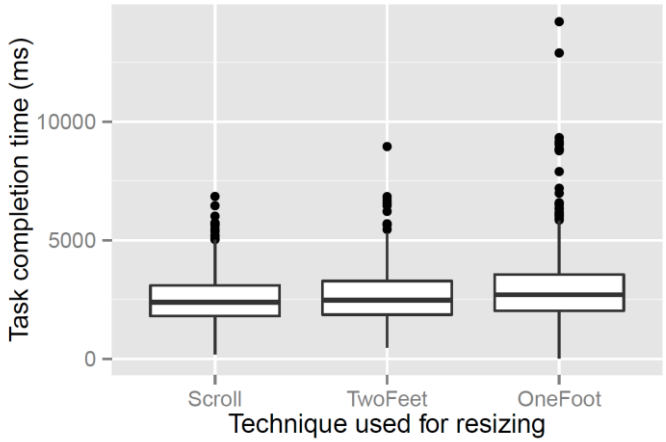


Figure 37 - Movement times for each technique in experiment 4.

We also computed the error rates for each technique: 0.15 for the scroll wheel, 0.089 for the two feet and 0.081 for the one foot (see Figure 38). We then compared the error rates for the conditions using a one-way repeated measures ANOVA. Mauchly's test indicated that the assumption of sphericity had not been violated, $W = 0.96, p = 0.86$. Results show a significant effect of the technique on the error rate, $F_{2,22} = 20.03, p < .05$. Bonferroni post-hoc tests showed that using the scroll wheel was significantly more accurate than the

other two conditions ($p < .05$), but no significant difference was found between using one or two feet ($p = 1.00$).

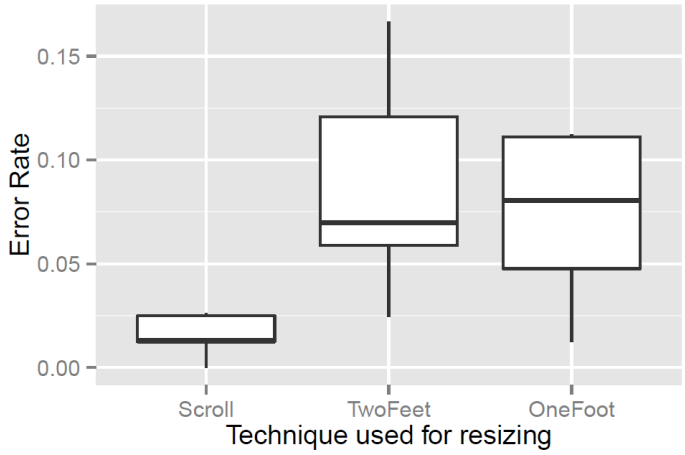


Figure 38 - Error rates of each condition in experiment 4.

The most preferred technique was the scroll wheel, chosen as the top technique by 85% of participants. Participants were divided between the feet techniques, with eight preferring two feet and five preferring one foot.

5.5.4 Discussion

We chose an increment value for each step of the scroll wheel so that users would not overshoot the thickness of the stroke of the target rectangle, but it also caused the scroll wheel to be slower, so it might have fared better with adjustments in its sensitivity. In terms of task completion times, the feet performed similarly to the hands, showing little overhead for the task being performed, but with a significant decrease in accuracy. Taking into account that users are more familiar with the scroll wheel and none of our participants had any experience with foot-operated interfaces, from these results we speculate that with training, the feet could match (if not outperform) the scroll wheel as a means of providing continuous input to applications.

Our results show that using two feet was significantly faster than using one. We suggest two explanations for this. First, because what mattered was the relative distance between the feet, users could place their feet wherever they felt most comfortable within the tracked area. Because in the one foot condition, what mattered was the absolute position of the foot, depending on how the user was seated, this position might not have been ideal, causing a decrease in performance. Second, as both conditions used the same calibration, moving two feet simultaneous would cause a twofold change in the size of the rectangle, as compared with moving just one foot, increasing the overall speed of the interaction. Despite being faster, almost 40% of participants still preferred one foot, citing that moving two feet was more tiring than moving just one.

5.6 Guidelines and Design Considerations for Continuous Input

Based on previous work, the quantitative and qualitative results from our experiments and our own experience while investigating the subject, we suggest a set of guidelines and considerations for designing desktop interactive systems that use feet movements as input.

Resolution: Our findings confirm the observations of Pakkanen and Raisamo that pointing with the feet should be limited to low fidelity tasks, in which accuracy is not crucial [195]. For example, when compared to using only the hands in experiment 4, the feet were significantly less accurate.

Visibility & Proprioception: In a desktop setting, the desk occludes the feet, which prevents direct input interfaces, such as the floor-projected menus in Augsten et al [7]. Moreover, foot gestures suffer from the same problems as other gestural interactions (see Norman [186] for a discussion of such problems), which are amplified by this lack of visibility of the limbs. Our second study showed that when designing such interactions, on-screen interfaces should provide a direct spatial representation of the movement of the feet. However, the lack of visibility of the feet is somewhat compensated by the user's proprioception: the inherent sense of the relative positioning of neighbouring parts of the body. Therefore, even though the user is not able to see their feet they still knows where they are in relation to their body.

Fatigue: Similarly to mid-air gestures, users report fatigue after extended periods of time using leg gestures. In all of our studies, participants reported that, in order to minimise fatigue, they preferred pivoting the foot around the heel to dragging the feet across the floor. Fatigue must also be taken into account when designing interactions where any foot is off the floor. In our experiment, when moving the feet across the floor, users preferred dragging the foot to hovering it over the floor.

Balance: Foot gestures performed whilst standing up only allow for one foot to be off the floor at the same time (except when jumping). While sitting down, the user is able to lift both feet from the floor at the same time, allowing for more complex gestures with both feet. To prevent from fatigue, such complex gestures should be limited in time and potentially also space. In this work, even though we tracked the feet in three dimensions, we only took into account their two-dimensional position in relation to the floor. It remains an open question how adding a third dimension could affect the interaction.

Chair & Spatial Constraints: The kind of chair where the user is seated may influence the movement of the feet. For example, the rotation of a swivel chair might help with moving the foot horizontally. Further, when both feet are off the floor, swivel chairs tend to rotate as the user moves which may hamper interaction. The form factor of desks, chairs and clutter under the desk also affect the area in which the user can perform gestures. This also offers opportunities for interaction, as physical aspects of the space can help guide the movement of the feet or serve as reference points. Another aspect that needs to be taken into account are the properties of the floor, which might influence the tracking (shiny floors will reflect the infra-red light emitted by the Kinect, creating additional noise) and interaction (floors covered in carpet or anti-slip coating may slow down feet movements, while smooth flooring may speed them up).

Rootedness: Mid-air gestures often suffer from the problem of *gesture delimiters*, similar to the classic *Midas touch* problem [127], as it is hard to tell specific actions and gestures from natural human movement [18]. This is less of a problem for feet gestures in a seated stance, because when sat down, most leg and foot movement consists of postural shifts, reducing the number of movements that might be recognised as false positives in gesture recognition systems. We addressed this problem in our studies by defining an area on the floor where the feet would provide input for the system, but in applications where it would be desirable to track the feet at all times, it is necessary to pay special attention to designing gesture delimiters that are not part of users' normal lower limb behaviour.

Footedness: The same way that people favour one hand they also favour one foot and, even though they are often correlated, there are exceptions to this rule, with approximately 5% of the population presenting crossed hand-foot preference [63]. Our findings indicate there is no significant difference between the dominant foot and the non-dominant one. These

results, however, reflect the performance of users with no experience with foot-based interfaces. It is not clear if this similarity in performance still holds for experienced users. Further, it is necessary to consider which foot will be used in the interaction, as crossing one foot over the other to reach targets on the opposite side might be too uncomfortable.

Hotspot: Touch-based interfaces, despite suffering from the phenomenon of `fat-fingers', can still provide a high resolution of input due to the small relative size of the contact area between the finger and the touch-sensitive area. Feet, however, provide a large area of contact with the floor. The designer can then opt for reducing the foot to a point or using the whole contact area as input. The former has the advantage of providing high resolution input, but users' perceptions of the specific point on the foot that should correspond to the cursor is not clear, as demonstrated by Augsten et al. [7]. Using the whole of the foot sole makes it easier to hit targets (as shown by Drury's modification of Fitts's Law [71]), but increases the chance of hitting wrong targets. Hence, if using this approach, the designer needs to leave enough space between targets as to prevent accidental activation.

5.7 Limitations

In this work, we described four experiments that attempt to characterise some fundamental aspects of the use of foot movements for interacting with desktop computers. These experiments, however, have some limitations. We collected data from a relatively small number of participants, so more precise estimates of the true value of the times and error rates presented here can certainly be achieved in experiments with larger pools of participants. Also, our participant pool was not gender-balanced in every study and did not cover a wide age range. We present results using only one tracking system that has several limitations of its own. For example, our prototype was implemented in *Matlab*, achieving a frame rate of 25fps, but the tracking speed could be improved by porting the system to a faster language, such as C++. While our results are in line with the ones in related work, further work is necessary to assess whether they translate to other foot interfaces.

5.8 Conclusion

In this work we took a bottom-up approach to characterising the use of foot gestures while seated. We implemented a foot tracking system that uses a Kinect mounted under a desk to track the users' feet and used it to investigate some fundamental characteristics of this kind of interaction in three experiments.

First, we presented ISO 9241-9 performance models for 1D and 2D foot pointing in a sitting position. Our results suggest little difference in performance between the dominant and non-dominant foot and that horizontal foot movements are easier to perform than vertical ones. We identified five strategies that participants employed to reach targets and found that the preferred one was rotating the foot around the heel. We also found that the biggest challenges for foot-based interaction in a desktop setting are difficulties in reaching small targets, hand-feet coordination, fatigue, friction with the floor, and overshooting targets. These findings are important because they help us complete our understanding of the potential of foot-operated interfaces and provide guidance for future research in this emerging domain.

Second, we studied the performance of each foot in controlling a single parameter in a unidimensional task. Our results showed no significant difference between the dominant and non-dominant foot, but it showed that horizontal movement on the floor is significantly faster than vertical. Also, users showed a preference for pivoting their feet rather than dragging them. Third, we looked at controlling two variables at once, comparing the use of one foot against the use of two (each foot using the same movement axis or different ones). Our results showed that the visual representation of the variables do matter, with the

performance for techniques that have a direct spatial mapping to the representation outperforming the others. It also showed that when the variables being manipulated are shown separately (such as in independent sliders), it is preferable to use two feet rather than one. Fourth, we analysed the use of the feet in parallel to the hands, showing that the feet perform similarly to the scroll wheel in terms of time, but worse in terms of accuracy, suggesting that with training and more accurate tracking systems, the feet could be used to support hand based interaction in a desktop setting.

Future work will focus on using these insights to design and implement techniques that can possibly enhance the interaction by supporting the hands in everyday computing tasks. While we provide some guidelines for design, it is still an open question as to which tasks can effectively be supported by the feet and the size of the cognitive overhead of adding such an interactive modality.

6 Multimodal Body Movement for 3D Interaction

"Think with your whole body."

Taisen Deshimaru

This chapter investigates our third and final proposition: that by treating body movements as multiple modalities, rather than a single one, we can enable novel user experiences. We propose breaking the sensing of the full body amongst multiple sensors, which together can provide more detailed tracking and enable up close interactions. We believe that multimodal interaction can offer several advantages, including improving the performance of current interaction techniques, enabling previously impossible interactions, and inspiring novel applications.

To investigate the topic, we chose the domain of 3D User Interfaces (3DUI) for several reasons. First, it offers a complex and multi-dimensional set of tasks that can highly benefit from the additional degrees of freedom provided by multimodal input. Second, recent technological advances, such as affordable stereoscopic displays and commercially available virtual reality headsets created a renewed interest in 3D interaction. Third, as the physical world around us is also three-dimensional, interaction techniques designed for 3DUIs can often be extended into other contexts, such as Ubiquitous and Mobile Computing.

The overall objective of this chapter is to explore full body multimodal interactions. Whereas sensors such as the Microsoft Kinect already offer full body sensing capabilities, they require a long distance between the user and the sensor. However, in most of our interactions with computing systems, we are quite close to them: sitting in front of our desktop computers, holding our mobile phones, standing in front of a public interactive kiosk, or lying in bed with our laptop over our legs. Whereas wearable tracking systems can offer full body tracking irrespective of the distance to the computer, they require user instrumentation, making it obtrusive and often uncomfortable.

To address this problem, we propose distributing the sensing of different body parts to multiple remote sensors. This approach offers several advantages. First, as each sensor has a specific body part to focus, they can be placed closer to the body, enabling up close interactions. Second, as a consequence of this, they can offer a higher sensing resolution. For example, whereas the Kinect cannot recognise finger gestures when tracking the whole body from far away, it can do so when placed near the hands. Third, by using remote sensors, we require no user instrumentation, allowing for more casual interactions.

In particular, we propose breaking body sensing into three. We track users' hands and fingers with a depth sensor placed above users' hands; users' eyes and head with an infra-red remote eye tracker; and users' feet with another depth camera facing down.

This chapter demonstrates how such a multi-sensor setup can benefit 3D interaction. We first investigate how the eyes can support mid-air gestures in 3D selection tasks. We conducted a user study that demonstrates how our multimodal approach provides a faster and more natural way of performing this task (Section 6.1). We then propose different foot-based interaction techniques to accomplish the canonical 3DUI tasks (Section 6.2). Finally, we bring it all together into a full body sensing platform in the form of an augmented arcade machine that enable public, playful and casual multimodal interactions (Section 6.3). Figure Figure 39 shows how the studies on foot interaction and gaze + gesture come together in enabling us to develop our full-body sensing system.

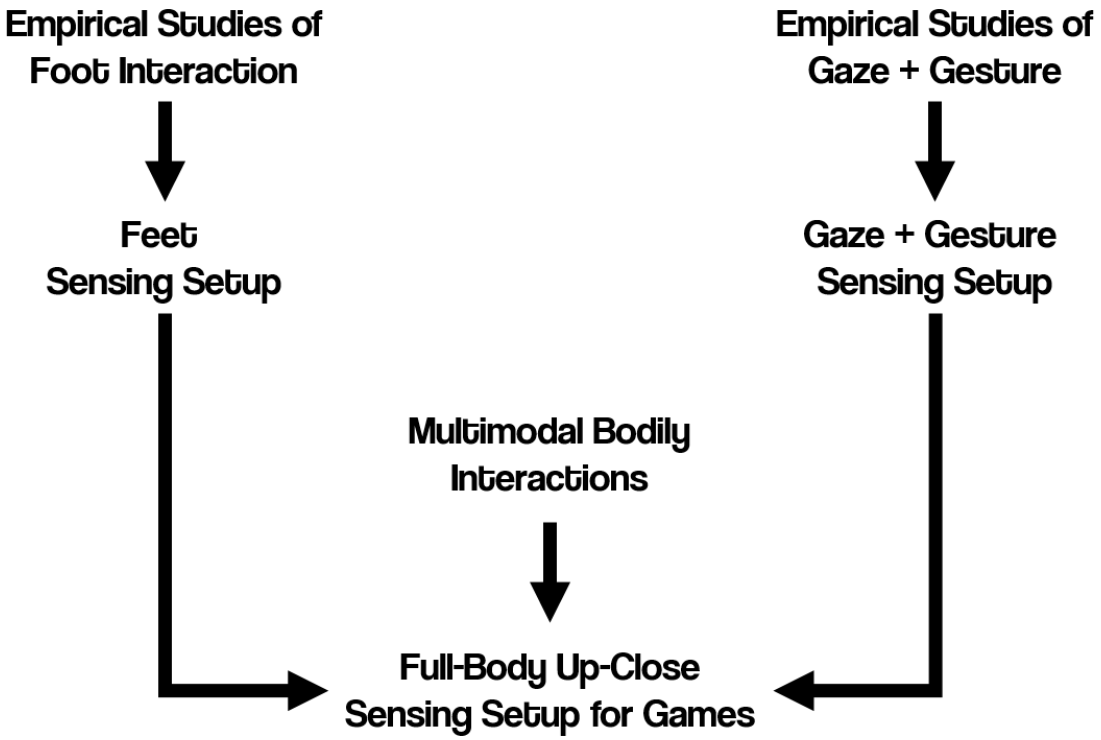


Figure 39 - We bring the findings and sensing setups from our empirical studies together into a full-body up-close sensing setup for games in the form of an arcade machine.

6.1 Gaze-Assisted Mid-Air 3D Selection

Interaction fidelity—the degree with which the actions used for a task in the UI correspond to the actions used for that task in the real world [30]—is an active topic of research in 3D

user interfaces (3DUI). Interfaces based upon free-space spatial input offer this fidelity for 3DUI due to their multiple degrees of freedom and high integration of dimensions of control (i.e. many degrees of freedom can be controlled simultaneously with a single movement) [29,112]. In particular, recent advances in unobtrusive motion capture (e.g. Kinect, Leap Motion) created a renewed interest in mid-air gestures for 3DUI.

In immersive virtual reality environments and on stereoscopic displays, such interactions allow users to manipulate virtual objects using interaction metaphors that relate more closely to real world interactions, for example, by using an isometric mapping between the virtual and physical spaces, users can reach virtual objects directly where they see them. However, a large number of 3D activities, such as gaming, graphic design and 3D modelling still rely on conventional monoscopic desktop displays. This setup creates a discontinuity between the physical and the virtual environments, and therefore do not allow users to directly grasp objects in three dimensions. In this desktop context, common mid-air interaction techniques for 3D selection are *Raycasting* (in which the user's hand controls a 2D point that determines the direction of pointing) and the *Virtual Hand* (in which the user controls a 3D representation of his hand and makes selections by intersecting it with virtual objects) [29]. See Argelaguet et al. for a survey of selection techniques for 3D interaction [5].

As the eyes provide a natural indication of the focus of the user's interest, eye trackers have been used for pointing in a wide variety of contexts without necessarily requiring a representation on the screen, showing higher speeds than conventional techniques [240]. Even though gaze pointing for computing input has been investigated since the 80's, studies on gaze pointing for 3DUI started with work by Koons et al., who built a multimodal interface integrating speech, gaze and hand gestures [145]. Early work was also conducted by Tanriverdi and Jacob, who found it to be faster than an arm-extension technique with a 6DOF magnetic tracker in a VR environment [261]. Cournia et al. found conflicting results that suggest gaze is slower than a hand-based Raycasting technique with a wand [54]. These works only investigated selection tasks, but in practice, common 3D interaction tasks involve further manipulation steps after selection, such as translation and rotation. Given that gaze alone is impractical for all steps, several works combined gaze with additional modalities, but few explored the context of 3D user interfaces. In particular, when using gaze for selection and mid-air gestures for manipulation, is there a cost in performance in switching modalities?

Even though gaze has been explored in a variety of multimodal configurations [255], few works explored the combination of gaze and mid-air gestures. Kosunen et al. reported preliminary results of a comparison between eye and mid-air hand pointing on large screen in a 2D task that indicate that pointing with the eyes is 29% faster and 2.5 times more accurate than mid-air pointing [146]. Hales et al. describe a system in which discrete hand gestures issued commands to objects in the environment selected by gaze [99]. Pouke et al. investigated the combination of gaze and mid-air gestures, but in the form of a 6DOF sensor device attached to the hand [211]. They compared their technique with touch, and found that the touch-based interaction was faster and more accurate.

The conflicting results of these papers highlight the importance of further investigating gaze selection for 3DUI, particularly considering that technical advances made eye tracking technology significantly more accurate, precise and robust than the devices and techniques used in previous works. In this work, we present an investigation of gaze selection for mid-air hand gestural manipulation of 3D rigid bodies in monoscopic displays. We conducted a study with three tasks. In the first task, we compared three 3D interaction techniques for selection and translation: a 2D cursor controlled by the hand based on Raycasting, a 3D cursor controlled by the hand analogous to a Virtual Hand and Gaze combined with mid-air gestures. In the second task, we also compared the same three techniques but in a selection and translation task involving multiple objects. In our pilot studies we found that when participants used the Gaze + Mid-Air Gestures technique, they reached out for objects even

though they did not have to. We hypothesised that this action was due to the clutching required for manipulation. To test this hypothesis, users performed a third task, in which we compared the selection time in the case where users were only required to select an object to the case where they also had to translate the object after selecting it.

Our results show that gaze selection is faster and more preferred than conventional mid-air selection techniques, particularly when users have to switch their focus between different objects. We also discovered a significant difference in the time to pinch after the object was gazed at between selection only tasks and selection followed by translation, indicating that the context of the selection impact the selection confirmation time.

6.1.1 Related Work

Human Prehension: Prehension is formally defined as "*the application of functionally effective forces by the hand to an object for a task, given numerous constraints*" [163], or more informally, as the act of grasping or seizing. Different authors proposed ways of modelling this process. In Arbib's model, the eyes (perceptual units), arms and hands (motor units) work together, but under distributed control to reach and grasp objects [4][163]. The perceptual schema uses the visual input provided by the eyes to locate the object and recognise its size and orientation. The motor schema can be divided into two stages: *reaching* (comprised of a quick ballistic movement followed by an adjustment phase to match the object's location) and *grasping* (including adjusting the finger and rotating the hand to match the object's size and orientation, followed by the actual grasping action). Paillard's model begins with the *foveal grasping*, in which the head and the eyes position themselves towards to object. Then, according to shape and positional cues, the arms and hands locate and identify the object, in open and closed loops, until finally grasping it, performing mechanical actions and sensory exploration [194][163].

In the context of mid-air gestures for 3D user interfaces, reaching is analogous to selection and grasping to the confirmation of the selection. In this work, we investigate how human prehension can be supported in a desktop monoscopic 3D environment. In all conditions we studied, grasping (confirmation) was performed by a pinch gesture, similar to how we would grasp physical objects, but the selection step varied across conditions. The 3D cursor includes a reaching step similar to normal prehension, only offset due to the discontinuity between the virtual and physical worlds. The 2D cursor also contains a reaching step, but only in two dimensions. The Gaze condition only requires foveal grasping, as when the user looks at the object, she only needs to pinch to confirm the selection. However, as we show in the results of task 3, when the user grasps the object for further manipulation, she still reaches out for it.

Mid-Air Interaction for 3D Manipulation: Due to our familiarity in manipulating physical objects with our hands, a considerable effort of the HCI community has been put into developing input devices and interaction techniques that leverage our natural manual dexterity to interact with digital content. An important interaction paradigm in 3D interaction is *isomorphism*: a strict, geometrical, one-to-one correspondence between hand motions in the physical and virtual worlds [29]. Even though isomorphic techniques are shown to be more natural, they suffer from the constraints of the input device (e.g. the tracking range of the device) and of human abilities (e.g. the reach of the arm) [30]. When targets are outside the user's arm reach, techniques such as Go-Go [212] and HOMER can be used to extend the length of the virtual arm [27].

Gaze in Multimodal Interactions: Gaze-based interaction is known to suffer from a few challenges [253]: inaccuracy (due to the jittery nature of eye movements and technological limitations), double-role of visual observation and control, and the Midas Touch problem

(the unintentional activation of functionality due to eye tracking being always-on [127]). To address these problems gaze is usually combined with other input modalities and devices.

Stellmach et al. investigated combinations of gaze with a wide variety of modalities [255], including a keyboard [254], tilt gestures [254][251], a mouse wheel [251], touch gestures [251][252][253] and foot pedals [92]. A common interaction paradigm in gaze-based interaction is that of *gaze-supported interaction*—gaze suggests and the other modality confirms [252]. An example of a gaze-supported interaction technique is MAGIC pointing, which warps the mouse cursor to the area around the gaze pointing [294]. Fine positioning and selection confirmation are performed normally with the mouse.

These works have shown that multimodal gaze-based techniques are intuitive and versatile enough to work in a wide variety of contexts, ranging from small mobile devices to large public displays [271].In this work, we have a similar goal to Stellmach and Dachselt, that of *seamless* selection and positioning [253]. Whereas in their work, they achieved this with different touch-based techniques for mobile devices, the context of 3D user interfaces requires extra degrees of freedom that are better suited for mid-air gestures.

Gaze and Mid-Air Gestures: Kosunen et al. reported preliminary results of a comparison between eye and mid-air hand pointing on large screen in a 2D task that indicates that pointing with the eyes is 29% faster and 2.5 times more accurate than mid-air pointing [146]. The techniques investigated in their paper are analogous to our 2D Cursor and our Gaze technique, as they also used a pinch gesture for selection confirmation. In this paper, we extend their work to 3D manipulation, also comparing them to a 3D cursor. Also, their task involved 2D translation of objects, whereas ours involves 3D translation.

Pouke et al. investigated the combination of gaze and mid-air gestures, but in the form of a 6DOF sensor device attached to the hand [211]. Their system supported tilt, grab/switch, shake and throw gestures. They compared their technique with a touch-based one, and found that the touch-based interaction was faster and more accurate, mainly due to accuracy issues with their custom-built eye tracker. We aimed to minimise tracker accuracy problems, by using a commercial eye tracker with a gaze estimation error of 0.4 degrees of visual angle. Our study also differs from theirs in that the mid-air gestures investigated by them were based on a tangible device, rather than hands-only gestures.

Yoo et al.'s system tracked the user's head orientation (as an approximation for the gaze point) and the 3D position of the hands for interacting with large displays [291]. Bowman et al. investigated pointing in the direction of gaze, but also approximating it to the head orientation [29]. Such approximations only work when the user is looking straight ahead. Hence they are only suitable for large scale interactions, such as with large displays and fully immersive virtual environments. In a desktop setting, the head orientation is not a good approximation for the point of regard, as the user is constantly facing the same direction.

Cha and Maier proposed a combination of gaze and mid-air gestures for a multi-display use case [43]. These authors presented architectural implementation details of their system, but did not present any evaluation results or interaction design decisions.

6.1.2 Experimental Setup

We recruited 12 right-handed participants (6M/6F), aged between 20 and 43 years (median=28). Three wore glasses and one wore contact lenses in the study. Figure 40 shows our experimental setup. Participants sat in front of an 18'' laptop running a custom application built in the Unity game engine. Gaze was tracked at 30fps using a *Tobii EyeX* tracker mounted under the display, with an average gaze estimation accuracy of 0.4 degrees of visual angle. Hands were tracked using an Asus *Xtion PRO LIVE* sensor, with resolution of 640x480 (30Hz) mounted facing down on a 0.82m×1.0m rig. Pose estimation and gesture recognition were performed using 3Gear Systems' *Nimble SDK*.

We implemented three interaction techniques for selecting and translating objects in our 3D scene:

- **Gaze (Gaze-Supported Mid-Air Gestures)**: the user looks at the object he wishes to select, pinch, move his hand to translate the object, and releases the pinch to disengage from the interaction.
- **2D Cursor (Raycasting)**: the user moves his hand on the plane parallel to the screen (up/down and left/right), which moved a cursor on the camera plane of the scene (moving the hand towards and away from the screen had no effect on the cursor). Targets were selected by hovering over them, (similar to a mouse cursor) and pinching. Then, the user moved his hand to translate the object and released the pinch to disengage from the interaction. Note that in this interaction technique, whereas the selection step uses only the XY coordinates of the hand, the translation step uses all three (XYZ).
- **3D Cursor (Virtual Hand)**: the user moves his hand around the space above the desk, which moved a sphere cursor in the virtual environment in three dimensions. Because we used an isomorphic mapping between the physical space and the 3D scene, any movement of the hand was directly translated in an equivalent movement of the cursor. To select an object, the user intersects the sphere cursor with the desired object and pinches. The user then moves his hand to translate the object and releases the pinch to disengage from the interaction.

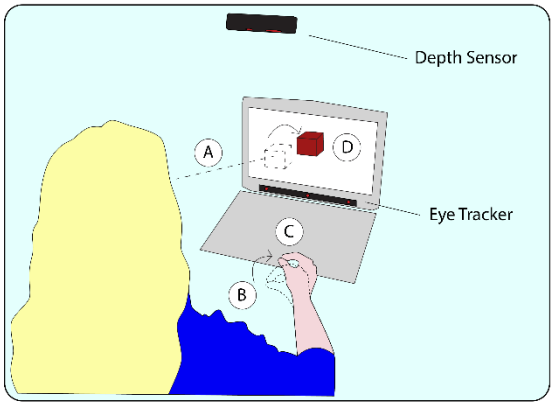


Figure 40 - Gaze selection for 3DUI: the user selects the object by looking at it (A), pinches (B), and moves her hand in free-space (C) to manipulate it (D).

Upon arrival, participants completed a consent form and a demographics questionnaire. We calibrated the eye and hand trackers with the manufacturers' default procedures. Participants then performed three 3D interaction tasks, described in the following sections. After all tasks were completed, we conducted an open-ended interview about their experience in using the interaction techniques.

6.1.3 Task 1: Translating a Single Object

6.1.3.1 Procedure

In Task 1, we compared completion times for two hand-based and one gaze-based selection techniques in a translation task. Participants were presented with a 3D environment containing one blue and one red cube (see Figure 41). The task was to pick up the blue cube with a pinch gesture, match its position to that of the red cube by moving their right hand whilst pinching, and drop it at the position of the red cube by releasing the pinch. When the blue cube intersected with the red cube, the red cube would turn green, indicating that the object could be released.

STUDY AT A GLANCE

Goal: Compare the task times for each step of a translation task using 3 interaction techniques

Method: Within-subjects experiment

Participants: 6M/6F (20-43y)

Dependent variables: Acquisition, Translation and Confirmation times

Independent variables: Interaction technique (Gaze, 2D and 3D cursor)

Results: Gaze is significantly faster

Participants performed the tasks in three blocks, each with 18 trials for each technique, in a counter-balanced order, for a total of 3 blocks × 3 techniques × 18 trials = 162 interactions. In each trial, the starting position of the cubes changed, but the distance between them remained constant. In the final block, after completing all trials for each technique, participants completed a questionnaire in which they rated each technique on a 7-point scale with respect to speed, accuracy, ease of learning and use, eye, hand and arm fatigue, intuitiveness, mental and physical effort, comfort, suitability for the task and personal preference. After completing all blocks they ranked the techniques in terms of speed, accuracy, comfort and personal preference. We discarded the first block from further analyses as a practice round.

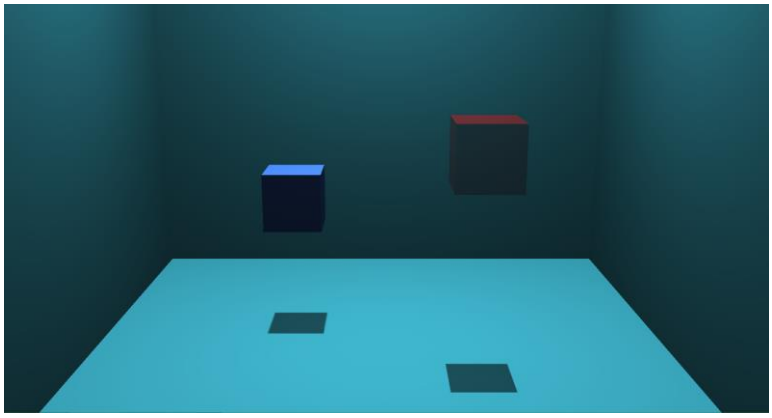


Figure 41 - Task 1: Users picked up the blue cube by looking at it and pinching. They then moved this cube until it touched the red cube, which, in turn, changed its colour to green.

6.1.3.2 Results

We compared the mean completion times between each technique across all trials, as well as the times of each step of the task, namely the time to acquire the blue cube (Acquisition), the time to pinch to confirm the selection (Confirmation) and the time to move it to the red cube (Translation). We tested the effects of the technique on the dependent variables using a one-way repeated-measures ANOVA (Greenhouse-Geisser corrected in case Mauchly's test revealed a violation of sphericity) and post-hoc pairwise t-tests (Bonferroni corrected).

The mean trial completion time using Gaze (3.76s) was 21.3% shorter than using the 2D Cursor (4.78s) and 37.0% shorter than using the 3D Cursor (5.97s) (see Figure 42). The effect of technique on mean completion time was significant ($F_{2,22} = 24.5, p < .01$) with significant differences between all combinations of techniques (p<.05).

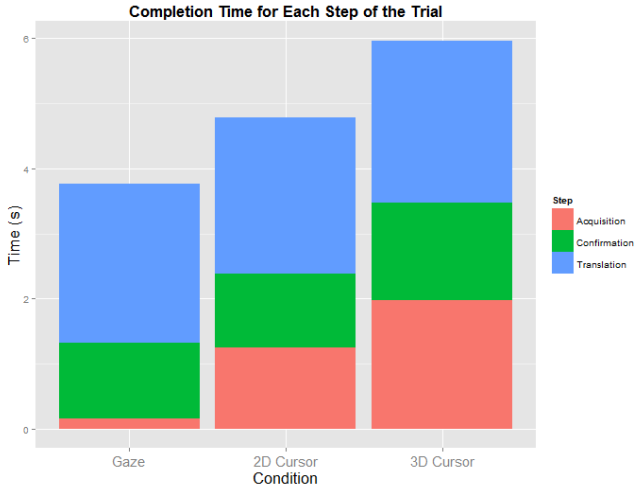


Figure 42 - Average task 1 completion time separated by step.

The Acquisition Time using Gaze (161ms) was 87.2% shorter than using the 2D Cursor (1.25s), and 91.9% shorter than using the 3D Cursor (1.98s), with a significant effect of the technique ($F_{2,22} = 194.5, p < .01$). Post-hoc tests showed significant differences between all combinations of techniques at p<.05. We did not find a significant effect of the technique neither on the confirmation time ($F_{1.2,13.2} = 3.1, p = .07$) nor on the translation time ($F_{2,22} = .12, p = 0.88$).

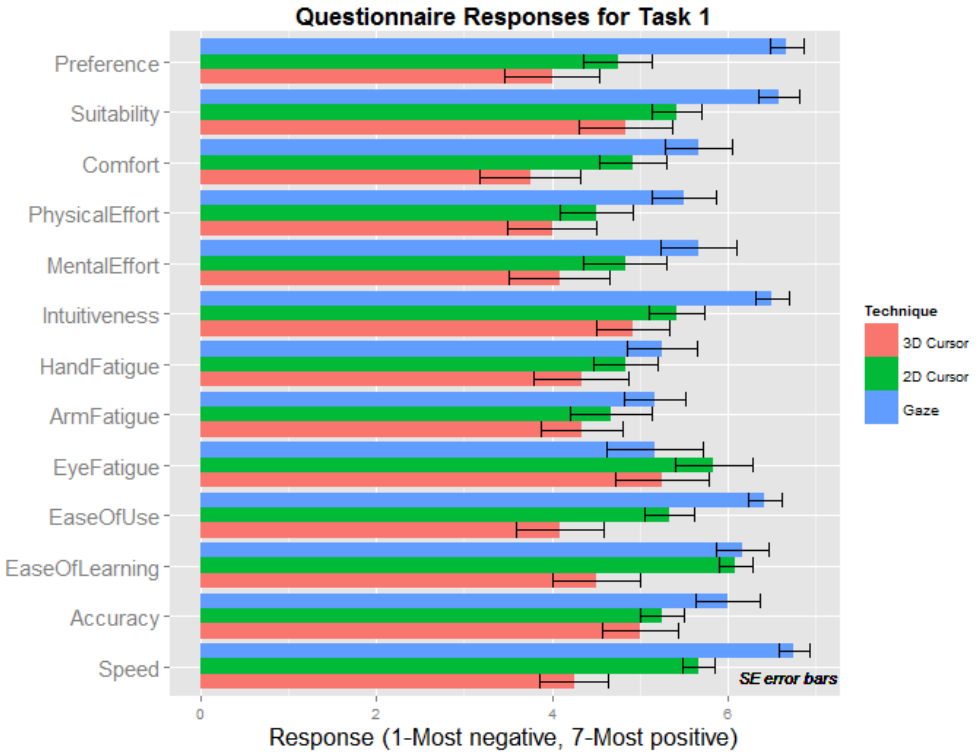


Figure 43 - Questionnaire after task 1. Gaze received higher scores than the other techniques along most dimensions.

In the questionnaires, Gaze received higher scores than the other two techniques along all dimensions (see Figure 33), except for eye fatigue, for which it scored the lowest of all three (but the difference was not statistically significant). Eleven participants ranked gaze as their preferred technique overall, with only one user preferring the 2D cursor. Nine users indicated the 3D cursor as the worst technique and three indicated the 2D cursor. A similar pattern was found for Accuracy, Speed and Comfort rankings (see Figure 44).

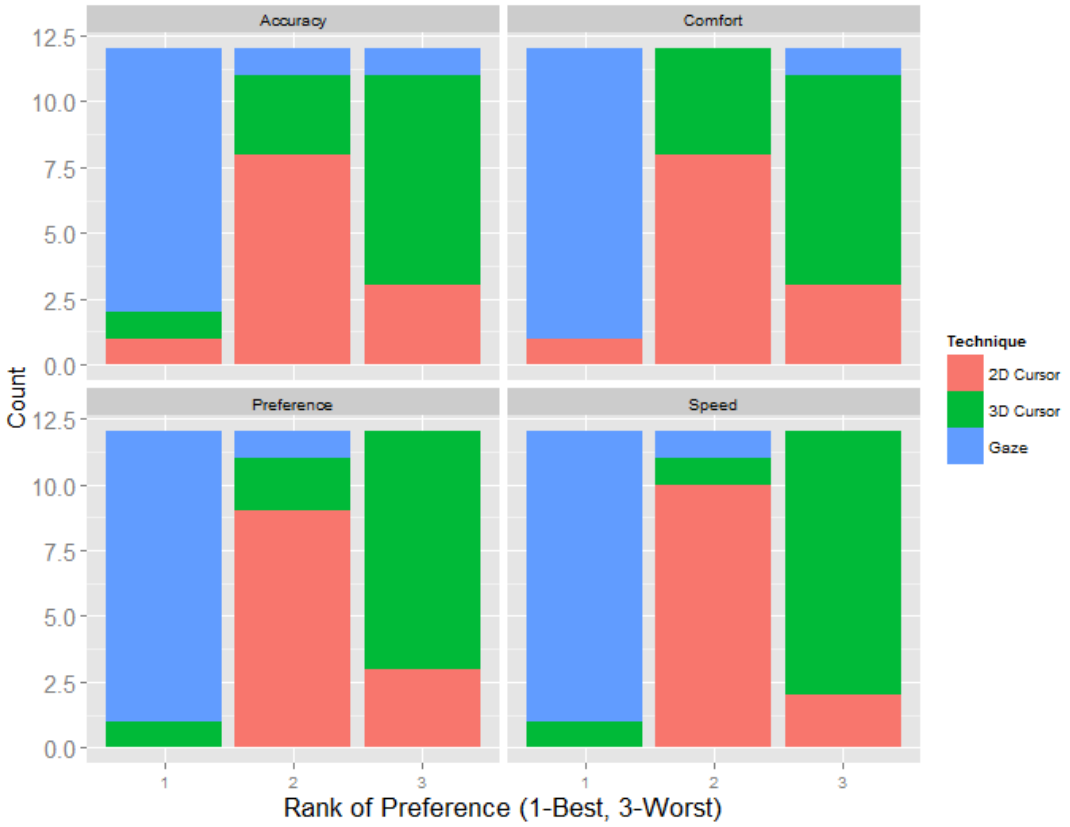


Figure 44 - Order of preferred techniques after task 1. Gaze was consistently the most preferred technique.

6.1.3.3 Discussion

The results from Task 1 are in line with Tanriverdi and Jacob [261]. Even though their setup was VR-based, it seems that Gaze also outperform other 3D selection techniques in monoscopic displays. Unlike Cournia et al., Gaze also outperformed Raycasting for selection, but as suggested by these authors, Raycasting performed better than Virtual Hand [54].

Both Tanriverdi and Jacob, and Cournia et al. investigated 3D selection, but not in the context of further manipulation. We also included a translation task to analyse whether the selection technique influenced the completion time of subsequent manipulation tasks (for example, by requiring clutching or adjustment of the hand position after selection). Because we found no significant difference in the confirmation and translation tasks, we cannot affirm that these interaction techniques have any effects on the manipulation task time, even though we observed certain hand clutching in the Gaze and 2D Cursor conditions. As shown in Figure 42, the only significant cause for the difference in the task completion time was in the object acquisition.

6.1.4 Task 2: Sorting Multiple Objects

6.1.4.1 Procedure

STUDY AT A GLANCE

Goal: Compare the interaction techniques in a cluttered environment

Method: Within-subjects experiment

Participants: 6M/6F (20-43y)

Dependent variables: Task completion time and error rate.

Independent variables: Interaction technique (Gaze, 2D and 3D cursor)

Results: Gaze is significantly faster, but more error-prone

In Task 1, we showed that the acquisition time using gaze is significantly shorter than using the other techniques. However, Gaze is known to suffer from inaccuracies, due to the jittery nature of eye movement, calibration issues and gaze estimation error. The goals of the second task were twofold: to investigate whether eye tracking inaccuracies would impair object selection in cluttered environments and to investigate how the faster selection times enabled by gaze can speed up tasks in which the user is required to rapidly manipulate different objects in sequence. We hypothesised that, because users do not have to necessarily move their hands to pick up new objects with Gaze, the fact that they could start the manipulation from wherever their hands were would speed up switching between objects.

Figure 45 - Task 2: Users picked up each chess piece and moved it to the appropriate side of the virtual environment.

Participants were presented with the same environment, now containing six black and six white chess pieces (see Figure 45). The right and left walls were coloured in white and black, respectively. Participants were asked to pick up each chess piece and move it to the appropriate wall. When the object collided with the corresponding wall, it disappeared. If the object collided with the wrong wall, it did not disappear, but we incremented an error counter. Each participant performed three blocks of 12 trials. In the last trial, after each technique, they answered the same questionnaire as before. After all blocks were completed, they completed the preference ranking questionnaire again.

6.1.4.2 Results

The mean time to put away each piece with Gaze (2.27s) was 38.4% shorter than with the 2D cursor (3.67s) and 49.4% shorter than the 3D cursor (4.47s) (see Figure 3-B). We found a significant effect of the technique on Completion Time ($F_{2,22} = 37.7, p < .01$) and significant differences between all combinations at $p < .05$.

The mean rate of incorrectly placed pieces with the 3D Cursor (1.92%) was 71.3% smaller than with the 2D cursor (6.70%) and 82.7% smaller than the Gaze (11.1%). We found a

significant effect on Error Rate ($F_{2,22} = 8.19, p < .01$). The post-hoc tests showed significant differences only between Gaze and the 3D Cursor ($p < .05$). No considerable differences were found in the questionnaire responses between the first and second task.

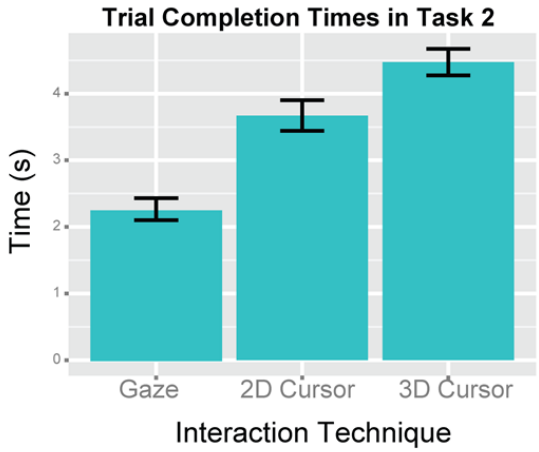


Figure 46 - Average trial completion times in task 2. Gaze was significantly faster than the other two techniques.

6.1.4.3 Discussion

The task completion times in Task 2 were significantly shorter than in Task 1. The reason for this is that whereas the selection step required precision, the translation step did not—as soon as the object hit the correct wall, the task was complete. For the hand-based tasks this represented a similar gain in speed (30% for the 2D Cursor and 34% for the 3D Cursor), but a much higher gain in speed for the gaze technique (66%). This shows that, even though it comes at a price of accuracy, Gaze is particularly well suited for tasks in which there is constant switching between different objects being manipulated. Examples of such tasks include organising furniture in architectural applications, playing speed-based 3D puzzle games and switching between different tools in a 3D modelling application.

6.1.5 Task 3: Selection Time

6.1.5.1 Procedure

In our pilot studies, we noticed an interesting phenomenon when observing participants using gaze-assisted mid-air gestures. In the Gaze condition, once participants looked at the object, they could pinch from wherever their hands were and start manipulating the object from there. However, users still slightly reached out to the general position of the object, either to open up space for subsequent manipulations or due to a natural tendency to reach out as when handling real objects.

STUDY AT A GLANCE

Goal: Compare selection times between a selection-only and a selection followed by manipulation scenarios.

Method: Within-subjects experiment

Participants: 6M/6F (20-43y)

Dependent variables: Time to pinch.

Independent variables: Task (Selection x Selection + Translation)

Results: Time to pinch is significantly longer when the user subsequently manipulates the object.

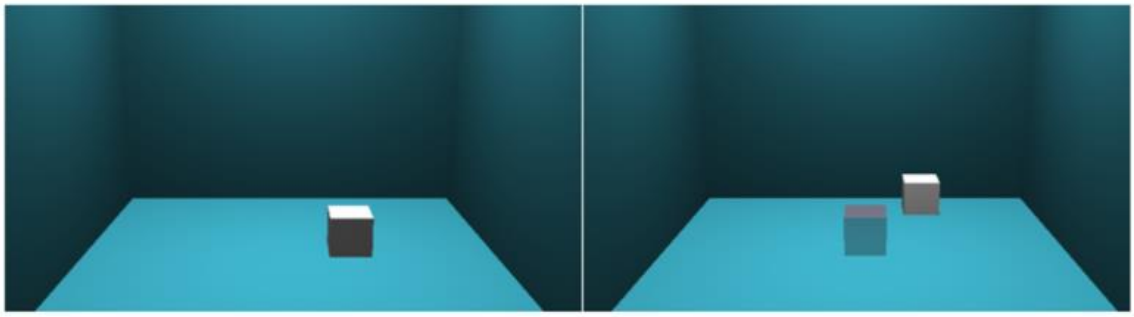


Figure 47 – Task 3: We compared the selection time between only selecting the object to the selection time with subsequent manipulation.

We hypothesized that this "clutching" before the translation would delay selection confirmation when compared to selecting the object without any subsequent manipulation. To test this, we conducted a third task with the same 3D environment (see Figure 47a). In one condition, a white cube appeared at random positions, but always at the same Y coordinate and at one of two Z coordinates (one in the foreground and one in the background). To reset the gaze point, between each trial the cube would show up at the centre of the environment. Participants were asked to look at the cube and make a pinch gesture, after which the cube would disappear. To avoid participants predicting the timing of the pinch gestures, we added a random delay uniformly distributed between 500 ms and 1.0 s before each trial. The second condition was similar to the first, but after pinching, the user was asked to drag the white cube to a red cube at centre of the environment (see Figure 47b). Participants performed three blocks, each containing 20 trials of each task (not counting the gaze-resetting steps), for a total of 3 blocks × 2 tasks × 20 trials = 120 interactions.

6.1.5.2 Results

We compared the time to perform the pinch gesture after having acquired the object with their gaze. The time in the Selection Only condition (557 ms) was 26.5% shorter than in the Selection + Translation (758 ms) (see Figure 48). A Welch's t-test revealed that this difference was significant ($t_{11} = -2.69, p < .05$).

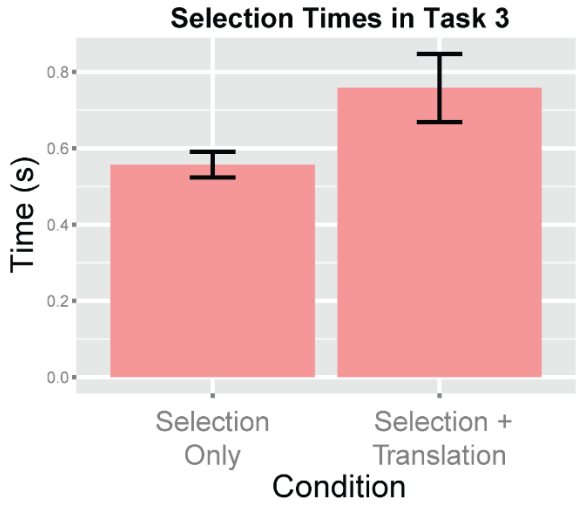


Figure 48 - Selection times in task 3. Users took longer to select the object when they were going to manipulate it afterwards.

6.1.5.3 Discussion

Our results show that the time taken to select an object with Gaze is significantly longer when the user plans on manipulating it afterwards. We offer three possible explanations for this phenomenon. First, when the user must translate the object after picking it, there is an additional planning step in the prehension process, adding some extra time for cognitive processing. Second, it is our natural behaviour to reach out in the general direction of where objects are. Third, with Gaze, even though the object can be selected from wherever the hand is, this initial position must allow enough room for the subsequent manipulation. Therefore, if the user's hand is not in an appropriate position, she must clutch it before picking the object up. From our observations, we believe the third explanation to be the most likely one.

6.1.6 Discussion

The results of Task 1 show that the acquisition time varied significantly between techniques, with Gaze being the fastest, followed by the 2D Cursor. We did not find a significant modality switch latency, as once the object was acquired, participants took approximately the same time to pick it up with a pinch gesture and move it to the target. As the results from Task 2 show, the advantage of Gaze is even stronger in tasks in which multiple objects are manipulated in sequence. This gain in speed comes at the cost of accuracy, particularly in densely populated environments. Participants' opinions on the techniques also confirmed that Gaze was the most popular technique.

In task 3, we discovered a significant difference in the time to pinch after the object was gazed at between selection only tasks and selection followed by translation. Although this difference is negligible for practical purposes, it reveals an interesting aspect of human behaviour when interacting using gaze. Even though the system was calibrated so that no clutching was necessary, participants still reached out in the general direction of where the object was positioned before pinching, similarly to how they would do with physical objects. Gaze selection elegantly supports this natural behaviour. This result suggests that gaze selection should be analysed in the context of the subsequent tasks, and not as an independent phenomenon.

We conducted our experiment in a desktop environment. The presented techniques could, however, be extended to standing interaction with large displays and immersive environments. Moreover, in stereoscopic displays, as the hands do not need to intersect the objects, Gaze-selection can be used without breaking the 3D illusion. Another limitation was that we only looked at translation tasks, but the same could be investigated for rotation and scaling.

Gaze-assisted mid-air manipulation allows users to select objects far away and manipulate them comfortably as if they were within reach. This allows users to rest their wrists on the desk, minimising the *Gorilla Arm* problem. This technique is are also particularly useful for monoscopic displays, where the inherent discontinuity between the virtual and the physical spaces do not allow for direct manipulation and often require an extra step for positioning the cursor on the target. In fact, participants reported not having to think about this step at all and that all they had to do was to think about the object and pinch, allowing for an arguably more immersive experience and an interaction with more fidelity.

6.1.7 Conclusion

In this work we evaluated gaze as a modality for object selection in combination with mid-air hand gestures for manipulation in 3DUI. Whereas previous works have found conflicting results on the performance of gaze for 3D interaction, we found that gaze outperforms other mid-air selection techniques and supports users' natural behaviours when reaching out for objects. Our findings suggest that gaze is a promising modality for 3D interaction and that it deserves further exploration in a wider variety of contexts. In particular, in future work we would like to explore how gaze can modulate the mapping between the physical and virtual environments, making it easier to reach distant objects, for example. Another avenue for investigation is how gaze can be incorporated into existing 3D applications.

6.2 Foot-Based Interaction Techniques for 3DUIs

The previous section investigated how gaze and mid-air gestures can be combined to create improved three-dimensional interactions. For that purpose we combined the hand and finger tracking capabilities of an overhead depth camera with the eyes and head tracking enabled by a remote eye tracker. In this section, we look at the other half of the body: the lower limbs. Inspired by the work presented in Chapters 4 and 5, we explored how foot movements can augment 3D interaction.

3D interaction tasks are inherently multi-dimensional, requiring highly expressive input devices capable of providing at least three degrees of freedom [107]. Conventional input devices such as keyboards and mice are not designed for this purpose, so previous research has explored new interfaces that take into account the user's 3D spatial context to facilitate interaction [28]. In this context, our feet can provide an additional interaction space. When their movements are suitably tracked, the input can be combined with the input expressed by our hands to form a more expressive interaction intent.

Technological advancements such as depth cameras and high precision touch-sensitive displays made mid-air and multitouch gestures a reality for a wide audience outside research labs. Whereas much work has explored the advantages of hand gestures for 3D interaction, feet gestures have been underexplored, especially in desktop settings. We aim to contribute in this domain through an exploratory investigation focused on the use of feet movements in the desktop setting. As a testbed environment, we built interaction techniques supporting 3D modelling tasks. Our work is motivated by the fact that even though gestural input can provide the necessary degrees of freedom for 3D interaction, they present challenges for stereoscopic displays, such as breaking the illusion of 3D [37]. This illusion is not hindered by foot interaction, as the feet are out of the user's field of view.

In traditional interaction users often have to switch the function of the mouse between several modes of operation, to cope with the high number of degrees of freedom required. For example, in a 3D modelling application, the user may move an object with the mouse on a 2D plane, use the mouse to rotate the camera and move it along the third axis, again with the mouse. In this example, by delegating camera control to the feet, users can enhance their performance by parallelising tasks that would normally be performed in sequence.

To investigate the possibilities of feet in 3D interaction, we used the foot tracker described in section 5.1. We then implemented four applications where the feet support each of the four fundamental 3DUI tasks [29]. To better understand the advantages and disadvantages of such modality, we conducted an exploratory user study with 7 participants where they interacted with our application and provided feedback in an interview.

This section contributes the design of four interaction techniques supporting fundamental 3D tasks, a discussion of users' feedback and guidelines for designing such interactions in the future.

6.2.1 Lower Body in 3D Interaction

The idea of using the feet to support three-dimensional manipulation was originally proposed by Choi and Ricci, who demonstrated an early application in which walking and leaning actions would rotate a cylinder in different axes [50]. Sangsuriyachot et al. controlled the rotation of a cube displayed on a tabletop with foot gestures tracked by a platform where the user stood [230]. Balakrishnan et al. used pedals and a foot mouse to select modes and control the camera in a 3D modelling application, while the hands manipulated a shape-sensitive tape that controlled an object on the screen [10]. This work builds on top of these by allowing for continuous and discrete input while freeing the feet from additional devices.

One of the primary functions of the feet in daily life is to support us while we walk, which makes navigation in virtual environments an intuitive application for them in HCI. Foot gestures that have been used for this purpose include leaning [62,154] and walking in place [185]. While we are interested in three-dimensional virtual environments, the focus of this work is on augmenting traditional desktop interaction, where it is not possible to lean or walk.

Previous work that evaluated the physical capabilities of the feet concluded that the feet can be from 1.6 [195] to 2 times [116] as slow as the hands, but this difference can be reduced with practice [88]. Researchers have demonstrated that the feet are suitable for tasks such as mode selection [238], non-accurate spatial tasks [195] and performing secondary tasks whilst the hands are busy [2]. This work aims to draw on these strengths to support the hands in manipulating three-dimensional interfaces.

6.2.2 Feet Support for the Canonical 3D Interaction Tasks

Bowman et al. suggest four main categories of 3D interaction tasks [28]: navigation, selection, manipulation and system control. For each category we implemented and evaluated a prototype demonstrating an interaction technique addressing it.

6.2.2.1 Navigation

In 3D modelling applications, users can manipulate the camera viewpoint by controlling the position and orientation of the camera. Common camera controls include dolly, roll, truck, orbit and pan. In conventional setups, this is usually achieved by holding a modifier key and

dragging the mouse. This forces the mouse to perform multiple functions, multiplexed over time.

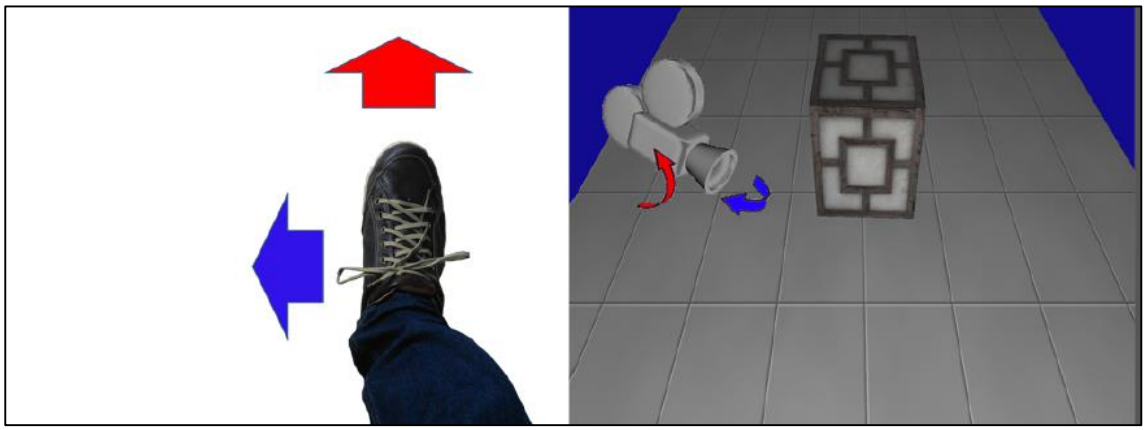


Figure 49 - Camera orbit with the feet. The horizontal offset (blue) changes the azimuthal angle and the vertical offset (red) changes the elevation angle around the current selected object.

In our implementation, the user moves the dominant foot to affect the camera's azimuthal and polar angles (see Figure 49). Foot tracking is relative to the starting foot location. The tracking is controlled by a key on the keyboard that toggles between an enabled and disabled state. Moving the foot causes the camera to orbit around the currently selected object. Horizontal movements affect the azimuthal angle, while vertical movements affect its polar angle. Using the foot to control the camera movement leaves the hands free to perform other tasks requiring higher precision. For example, positioning an object so that it matches to visual features in the environment that need to be observed from multiple orientations, e.g. an architectural application where the user needs to place a tree so that it does not occlude house windows. Foot-based camera manipulation can support these tasks by allowing users not to interrupt their interaction flow, performing in parallel tasks that would have been performed in sequence.

During pilot testing we observed that mapping camera XY trucking/dolling to the non-dominant foot can easily lead to fatigue. Two mapping choices are possible: relative or absolute. With relative mappings, users would need to perform multiple forward/backward or right/left movements with their foot: a first movement causes the camera to advance; a second movement is needed to return to the starting position (e.g. by not tracking it while it is lifted) and perform another movement. Depending on the size of the environment and the rate of change associated to this movement, traversing it would lead to fatigue as multiple back and forth movements would need to be performed.

Absolute mappings, i.e. mapping the feet-tracked area to the whole environment so that moving the foot between the boundaries would result in placing the camera in the corresponding location. For this to be feasible, users would need to see an on-screen world-in-miniature representation of the environment. By visualizing the position of their non-dominant foot through a pointer in this representation, users would have a better understanding of the location they are about to move the camera to. We believe this approach to have more potential and we will explore it in future work.

6.2.2.2 Object Manipulation

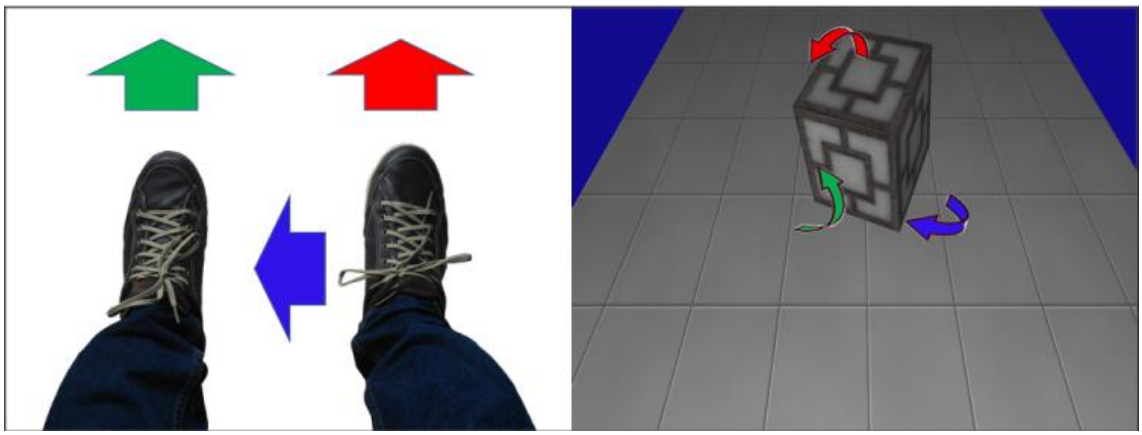


Figure 50 - Object rotation with the feet. The horizontal and vertical offsets of the right foot affect the object's yaw (blue) and pitch (respectively). The vertical offset of the left foot affects its roll (green).

We wanted to explore whether feet-based mappings could be used to provide a way to manipulate an object while the user focuses his hands on tasks requiring higher precision. Analogously to the previous application, we mapped yaw and pitch to horizontal and vertical movements of the dominant foot, respectively (see Figure 50). In addition, roll was controlled by using horizontal movements by the non-dominant foot. We chose to map rotation to the feet as similar considerations as those made in the previous example can be applied on their use on manipulations affecting object translation. This can be combined with mouse and the keyboard to provide a full 6DOF manipulation alternative to specialist input devices, so that the hands control translation while feet are mapped to rotation.

6.2.2.3 Selection

Foot input can also be used to express discrete inputs by means of foot gestures, such as a swipe. A foot-based swipe gesture is a rapid gesture performed by the user either to the right or to the left of the starting position (see Figure 51). Foot swipes can also be used vertically. The gesture is detected by analysing the velocity of the movement and comparing it to an empirically determined threshold. A cool-down period of 250ms avoids unintended multiple gestures. In chapter 5, we determined that lateral movements are easier than frontal ones. Indeed, horizontal foot swipes can be achieved by keeping the foot still and pivoting it left and right, whereas vertical swipes require the user to drag the foot forward or backward. In our implementation, horizontal swipes allow users to switch the currently selected object to the one whose 2D screen projection is to the right or to the left of the starting one.

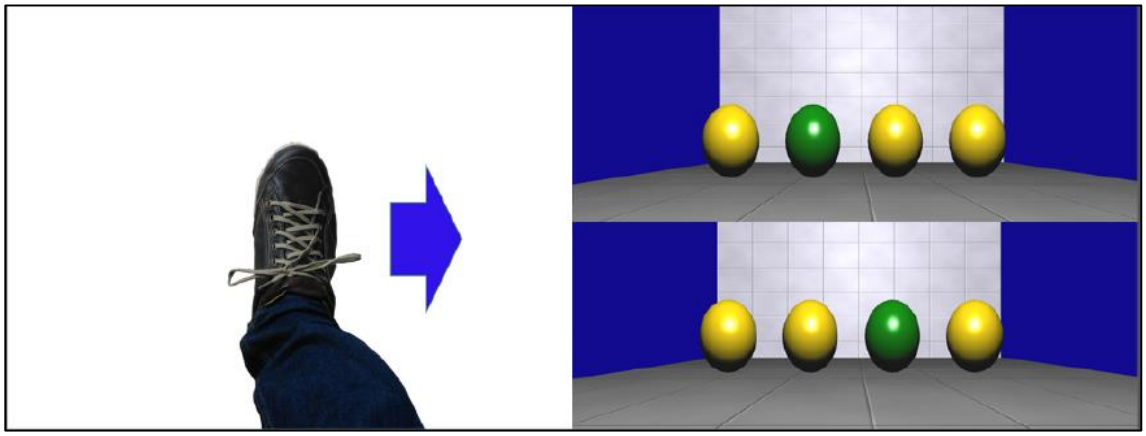


Figure 51 - Object selection with the feet. The user iterates over objects by performing a swiping gesture either to the right or to the left.

This technique can be used in a wide range of scenarios where users would normally have to switch between a manipulation and a selection mode. In a modelling application, users need to move from object to object to perform different operations on them. Foot swipes provide a convenient alternative for changing the currently selected object, without breaking the interaction flow. Foot-swipes can also be implemented to enable semantic selection within an object (i.e. selecting vertices, edges or faces) by mapping each foot to one of these two selection operations.

6.2.2.4 System Control

Finally, we designed a "radial foot menu" divided into four quadrants, each associated to a different colour parameter (e.g.: hue, saturation, brightness and alpha, see Figure 52). The dominant foot controlled a slider affecting the currently selected function from the menu, while the non-dominant foot selects the property to be changed.

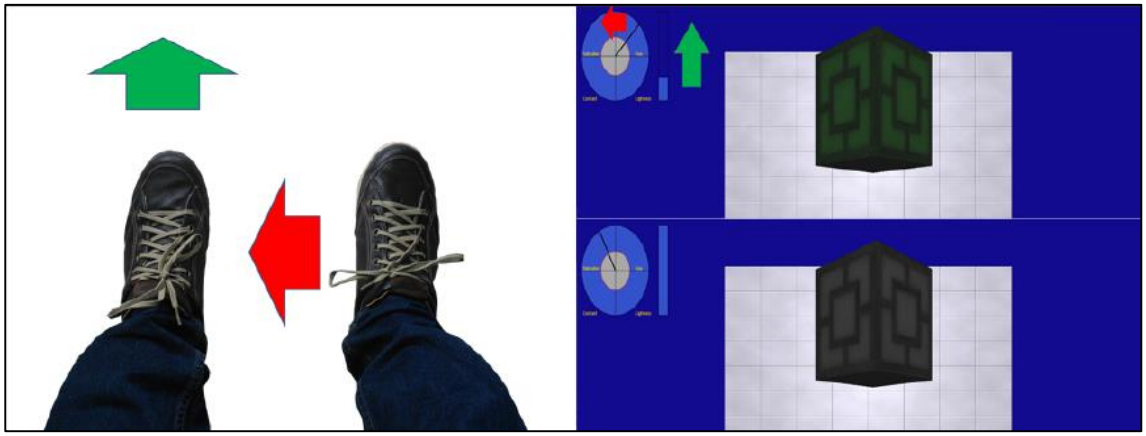


Figure 52 - Radial foot menu. The position of the right foot selects the parameter to be controlled and the vertical offset of the left foot selects the value, in this case, the saturation of the object's texture.

Adjusting material properties is a task that requires the user to visualise the effect of the changes on the model it is applied to. If further changes are necessary, the user has to return to the editing environment and make adjustments. Foot-based movements can support this task by allowing users to visualise the effect that parameter adjustments can have on the model they are working on. In this way, the hands are free to manipulate the object and

observe it from different viewpoints while continuing to make adjustments. Once satisfied with the results, a modifier fey can disable foot-tracking and finalise the operation.

6.2.3 User Study

STUDY AT A GLANCE

Goal: Validate foot-based 3D interaction techniques

Method: Exploratory user study

Participants: 7 (3M/4F)

Procedure: Use foot techniques to accomplish the 4 canonical 3D tasks and provide qualitative feedback.

6.2.3.1 Participants & Apparatus

To better understand the advantages and disadvantages of such modality and to get user feedback on the interactions it affords, we conducted an informal user study with 7 participants (3M/4F), with ages varying from 19 to 30 (mean 26.3). All participants had prior experiences with 3D movies; five reported to regularly drive a car; none had any prior experience with a foot mouse or similarly foot-operated device. Our experimental setup (see Figure 53) used an Asus VG278h 3D monitor to display the applications. The stereoscopic effect was obtained using DirectX 11.1's native support by computing the stereo pairs, without resorting to an automatic driver implementation. We used a custom 3D engine using SharpDX's DirectX11 C# port.

Figure 53 - Experimental setup: users performed the 4 canonical 3D interaction tasks displayed on a stereoscopic display with the feet tracked by a Kinect sensor

6.2.3.2 Procedure

Participants were asked to use the four applications previously described and to provide feedback in a questionnaire and in a structured interview. In the camera manipulation task, they were asked to rotate the camera to read a number written on a hidden side of a cube; in the object manipulation task, users were asked to rotate a cube in its 3 axes so that the side containing the number was visible and upright (the orientation of the cube was reset so that the number was randomly hidden again); in the selection task, users were asked to select specific objects by foot-swiping left and right to iterate over a group of spheres in a 3D environment; in the system control task, users were asked to select the saturation property in a radial menu and change its value in order to make the selected object greyscale. After each task, we conducted an interview and participants filled in a Likert scale questionnaire. The experiment took approximately half an hour for each participant.

6.2.3.3 Results

Participants were able to quickly grasp how to operate the four different techniques. They reported low signs of frustration (M = 2.0; SD = 1.05, from 1 – very low to 5 – very high) and a low cognitive demand (M = 1.93; SD = 1.12). When asked whether they would be to coordinate their hands with one or both feet, they all replied positively (M = 4.39; SD = 0.69), with some distinctions. Two participants shared the opinion that feet are better used for coarse control tasks such as camera manipulation, whereas object rotation was deemed to require more precision. Thus, using feet-based movements for secondary tasks appears a more promising direction, as this leaves the hands free for fine-grained manipulations.

Participants observed that they were able to use their feet so to get the desired results. They confirmed our assumptions that pivoting is more comfortable than sliding one's foot across the floor. The floor surface used in our experiment consisted of carpet which, as noted by participant #1, introduced some friction. The main issues we observed in this regard concerned the difficulty in estimating whether feet were still in the tracked area and how to avoid unwanted movements, which we address in the next section.

On the suitability of foot-movements for 3D interaction, participant #1 said: "[They are] good for continuous 3D interactions that don't have to be perfect, you can do many different things at the same time".

Participant #6 thought that: "Camera manipulation felt easier than rotating the object, as I think that it is better when the object that you are working on is fixed; object rotation might be easier to perform by having each foot control one axis".

On the ease of use and comfort, participant #2 said: "Foot movements might help the health of the user, as it adds activity during computer work".

Participant #4 stated that: "Simpler feet movements are easier to perform, so I would be able to coordinate [foot movements] better with hands movement".

Participant #1 expressed some concerns about unwanted movements: "If I move my feet accidentally, and this is something I often do, it would probably trigger some unwanted stuff. Also, if I have to move too much with my legs, it might be too exhausting".

Participant #2 stated that: "Rotation should not require the user to twist the foot too much".

6.2.4 Discussion

During our study, we collected a number of observations and suggestions that we discuss in the following paragraphs.

Limitations. In our implementation, feet tracking could be enabled or disabled by pressing an associated key. Once enabled, participants explored the interaction techniques without reverting to a non-tracked state. This introduced a form of the "Midas Touch" problem, where participants accidentally performed unintentional movements which affected the state of the system, causing some frustration. In a real application scenario, users need an unobtrusive way to enable and disable the tracking. The keyboard key is one of the options we have explored.

An alternative approach consists in dividing the floor between tracking and rest areas. Only when feet enter in the tracked area, the system is affected. Further options consist in using a foot-tapping gesture (i.e. lifting one's toes and tapping the floor) to toggle between tracked and untracked states. The microphone in the Kinect can be used to detect the audible impact and support the detection of the gesture. Another alternative consists in raising one's toe and pivoting the foot to enable tracking and, at the same time, affect the environment.

Estimating the tracked area's boundaries was another of the issues we observed. Participants were not aware of where exactly these boundaries were. We believe that this

can be addressed by displaying a warning icon when the user is in proximity of the boundaries of the tracked area. Another issue we noticed is that, due to limitations in the depth-sensing technology, it became hard to distinguish when both feet were close together.

Design Guidelines. As we have previously highlighted, foot-based UIs can play an effective role in supporting primary tasks by delegating secondary ones to the feet. Users can use their hands to perform fine-grained manipulations with other input devices while feet can be used to control those aspects that would otherwise require frequent mode switching (i.e. between manipulation and system-control modalities). To avoid confusion, applications should inform users when tracking is active and provide the option to undo feet actions directly through feet gestures.

Hand-Foot Coordination. Based on our observations, we believe that in order to minimise users' cognitive burden, each modality should be assigned to actions that can be performed independently, e.g. assigning hands to manipulation and feet to camera rotation. Mappings that require integration of DOFs split across modalities should be avoided. The extents to which foot and hands are able to work in parallel on different aspects of a common goal will need to be evaluated in future studies. Visualizing feedback of how feet are affecting the system might improve coordination.

Fatigue. In order to maximise the potential of this novel approach to 3D interaction, we believe it is important to focus on the design of foot mappings that are comfortable to the users. We have observed how horizontal foot movements were preferred by participants, whereas vertical foot movements were deemed to have higher potential for causing fatigue. In addition, pivoting movements were preferred over sliding movements. Thus, it appears that in-place movements hold more potential than dragging one's feet across the floor. Friction hinders the latter, while pivoting appears to be easier to control and requires less effort. Our observations indicate that short foot motions, in terms of amplitude and actual movement, were the most comfortable and less fatiguing ones.

6.2.5 Conclusion

In this section we have presented an exploration of feet-based interaction techniques supporting 3D tasks. Through an informal user study we determined that foot-movements are easy to learn and can be used to perform 3D manipulations and control the system. We described how it is possible to address the limitations emerged during the study.

6.3 Multimodal Full-Body Sensing in an Arcade Machine

In the previous two sections, we demonstrated novel ways in which the upper body (via mid-air gestures and gaze) and the lower body (via free-form feet movements) can be used in 3D user interfaces. We believe that a subdomain of 3DUI that can strongly benefit from this multimodal setup is gaming. However, we recognise that the complexity of sensing setup makes it unlikely that players will have access to this kind of interactions at home. To address this problem we propose incorporating all the different sensors in to a single unified platform, in the form of an augmented arcade machine.

6.3.1 Introduction

> *"Absolutely, it can come back. Creativity will bring anything back. There's so much technology out there that can't be packaged in the home environment."*

Nolan Bushnell, founder of Atari and arcade pioneer, when asked whether arcade machines can come back [133]

Arcade machines represent an important chapter in the history of video games. At a time when computers were too bulky and expensive for household use, people of all ages could get together at specialty shops and play the addictive coin-operated games. The arcade phenomenon peaked in the late 70's and early 80's, an era dubbed The Golden Age of arcade games. The advent of the home console brought an end to this era in the West, though arcades are still relatively popular in emerging Asian markets to this day.

Figure 54 - Arcade+: Input is captured via combinations of a joystick, buttons, eye tracking, feet tracking, hand tracking and touch gestures

Despite their declining popularity, arcade machines enable players to have access to input devices that they might not have at home. In this section, we propose bringing back the form factor of arcade machines as a means of deploying and evaluating multi-modal games. We designed and built *Arcade+* (see Figure 54), an arcade machine augmented with modern input devices to deploy and evaluate novel multi-modal interaction techniques for gaming. Our system offers gaze, mid-air gestures, feet and multi-touch sensing as well as conventional arcade joystick and buttons. Its modular design allows for adding and replacing input devices. We also make our designs available for the community to build upon and extend.

The contributions of this work are: (1) a sensing platform that can simultaneously track players' eyes, finger gestures, touch, and feet at close range; (2) three games that demonstrate how the availability of a whole body sensing setup can inspire novel game mechanics; (3) the schematics and instructions to replicate our design.

6.3.2 Full Body Expressions in Gaming

Body movements are progressively gaining a more central role in gaming. Most modern consoles enable some form of motion input, ranging from simple accelerometers in mobile phones to full body tracking using depth cameras. Bianchi-Berthouze argued that the reason for this is that body movement "*affects cognitive and emotional processes*" and "*the increased involvement of body movement during game plays results in increased enjoyment*" [21].

Bianchi-Berthouze et al. suggested that full-body movement facilitate the immersion in virtual environments, enables affective elements of human communication and unleashes regulatory properties of emotion, finding statistically significant relations between body movement and engagement [20].

Players' movements depend on their goals and setting. In games that require full body movements, when the goal of the player is to win, they tend to use small, controlled movements, whereas when their goal is to relax, they tend to use more realistic moves [205]. In a desktop setting, whereas engaged players often sit still to help focus their attention on the game, disengaged players tend to fidget, lean back and yawn [19].

In a gaming context, Bianchi-Berthouze distinguished five types of body movement: *task-control movements* (those set by game designers to control the game), *task-facilitating movements* (to assist the player in distributing cognition over body resources), *role-related movements* (related to the role adopted by the player in the game scenario), *affective expressions* (that reflect the affective state of the player) and *expressions of social behaviour* (that facilitate and support interaction between players) [21].

Whole body movements have also been used in Exertion Interfaces for games. An Exertion Interface is one that "*deliberately requires intense physical effort*" and they have been widely used for gaming, especially for sports games [180].

6.3.3 The History of Arcade Machines

The origins of Arcade machines can be traced back to carnival games, such as target shooting and fishing for prizes. David Gottlieb's *Baffle Ball* (1931) was the first pinball machine to replicate this pay-per-play business model with a coin slot. For an inspiring catalogue of vintage arcade machines, see Ford [85]. These machines were often built by the same manufacturers of gambling machines. This raised substantial controversy because of the connection between gambling and organised crime, to the extreme of New York banning pinball machines from 1942 until 1976. *Galaxy Game* (1971), built at Stanford University, was the first coin-operated video game and was soon followed by *Computer Space* (1971), the first mass-produced video arcade game. *Computer Space* received a lukewarm response, as it proved too complex next to gambling machines in bars and too confusing next to pinball machines in arcade shops [69]. The arcade business really showed its potential with the release of *Pong* (1972), which, with its simple rules—'*avoid missing ball for high score*'—became the first commercially successful coin-operated video game, selling over 35,000 units. The Golden Age of arcades in the late 70's and early 80's yielded several commercial hits coming from all over the world, especially from Japan. *Space Invaders* (1978), *Asteroids* (1979), *Pac-Man* (1980), *Donkey Kong* (1981), among others, sold hundreds of thousands of units and created popular icons that remain relevant even today.

Several reasons contributed to the end of this era [69]. First, to make it a profitable business, arcade games relied on ramping up the difficulty in order to make the player lose and allow another user to try it, which naturally frustrated many players. As home consoles entered the market, their business model allowed for games with more depth and longer play times. Second, the popularity of arcade machines generated a large number of similar games being made, effectively saturating the market. Third, a moral crusade led by concerned parents resulted in lawsuits against video game arcades leading to restrictions and even bans across the U.S.

The late 90's and 2000's brought a renewed interest in arcades with the release of games that explored novel ways of interacting with them. The rich and specific form factor of these machines allowed for input devices that created new experiences that home consoles simply could not provide. *Time Crisis* (1996) and its light gun, *Dance Dance Revolution* (1999) and its dance pads, *Mario Kart* (2005) and its camera, all became successful examples of the additional value that arcade machines could provide over home consoles. In this work, we propose using the classic form factor of Arcade machines as a tool for research on modern multi-modal interaction techniques for gaming.

6.3.4 Arcade+

The design of Arcade+ was guided by three principles. First, we wanted to remain true to the classic form factor of 80's machines. Second, we wanted to enable novel user experiences by using the whole body as a game controller. Third, we wanted to provide a reproducible and extensible design for the community to replicate and improve.

Figure 55 - Arcade+ Components: (a) Asus Xtion PRO Live; (b) Tobii EyeX; (c) Kinect for Windows; (d) Pioneer speakers; (e) Iiyama Multitouch Display; (f) Joystick and Buttons

Figure 55 shows our arcade machine, with dimensions of 1.95m×0.92m×0.79m. It is supported by an aluminium structure made of 45x45L profiles and its cabinet is made of plywood. It is powered by a Windows PC with an Intel i7 Quad Core 3.60GHz CPU, 16GB of memory, a GTX 970 Nvidia *Geforce* graphics card. For output, we included an Iyama *TF2234MC-B1X* frameless 22'' touch-enabled screen, which is scratch- and water-proof. We also installed a 190W set of Pioneer car speakers and amplifier for audio output.

For input, we included traditional arcade buttons and a joystick on a removable panel attached to the front of the machine that communicate with the PC through an Arduino UNO. By making the dashboard removable, we can create different modular designs that can be switched according to the game. For example, we can have one with another display for dual-screen games or another with a set of light pistols for rail shooters. We included a Tobii *EyeX* tracker installed below the screen to allow for gaze-enabled games. The tracker is mounted on a tilting platform to allow users to adjust it to their height. We also added a button next to the eye tracker to trigger the calibration procedure. An Asus *Xtion PRO Live* installed above players' hands allows mid-air gestural interaction. Hand pose recognition is performed using 3Gear's *Nimble SDK*, which is able to track precise finger gestures. Additionally, we installed a Microsoft *Kinect for Windows* under the dashboard to track the feet, using the foot tracker developed by Simeone et al. [242]. The system was designed in Autodesk *Inventor* and we made the source files for the structure and casing available online at http://eyecardeproject.blogspot.de. The website also contains the hardware and software specifications of our system, as well as a guide on how to replicate it.

6.3.5 Games

To illustrate the capabilities of the system in enabling multi-modal games and inspiring new game mechanics, we implemented three games: *Feyerball Mage* (see Figure 56a), *StarGazing* (see Figure 56b) and *Virus Hunt* (see Figure 56c). All games were built with the *Unity* game engine. The idea behind the games was to let the different combinations of modalities inspire new mechanics. The three games respectively combine full body and gaze; arcade controls and gaze; and touch and gaze.

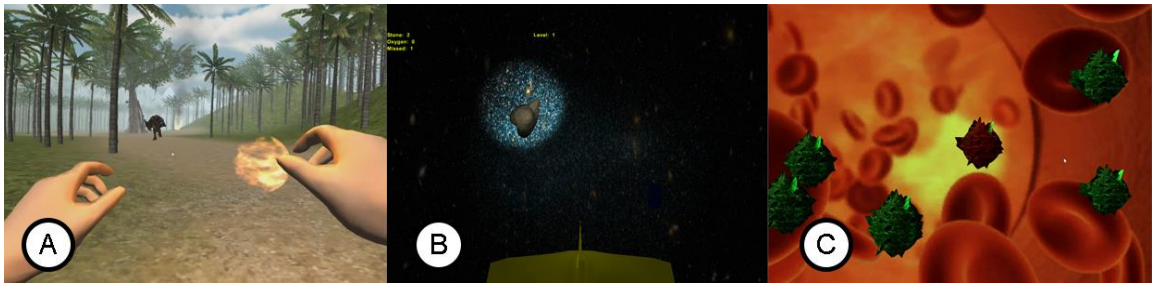


Figure 56 - (A) Feyerball Mage: throw fireballs with mid-air gestures at the direction of the gaze point, and navigate with foot gestures; (B) StarGazing: gaze reveals asteroids that can be shot at with the arcade buttons; (C) Virus Hunt: touch the viruses to kill them, but avoid looking directly at them.

6.3.5.1 Feyerball Mage

Aim | Turn / Fireball | Throw | Walk

Feyerball Mage is a first-person game (see Figure 56a), in which the player controls the title character. The game explores the play concept of combining full body interaction with gaze. The goal is to find all pieces of an ancient artefact that are scattered around an island, populated by trolls. The game tracks players' hands (rendered on the screen) and eyes. By making a pinch gesture and releasing it, the mage throws a fireball in the direction of the gaze point. Users can walk forward by stepping in place and turn around by moving their hands towards the edge of the screen. This game contributes the first exploration of how gaze tracking can be combined with full body tracking in a game environment.

6.3.5.2 StarGazing

Spotlight / Aim | Move | Shoot

In *StarGazing* (see Figure 56b), players must protect their spaceship from incoming asteroids. However, the darkness of space prevents them from seeing the asteroids, unless players aim the spotlight controlled by the eyes. Therefore, players must scan the environment with their eyes searching for asteroids and shoot them with the arcade buttons to destroy them. *StarGazing* was designed to explore the play concept of limiting the field of view to reduce the effectiveness of saccades. Saccades are the rapid movements made by our eyes when switching between fixations. With such movements, users can quickly switch

between looking at different targets on the screen. By limiting the field of view, we effectively remove players' peripheral vision within the game environment, forcing users to scan the environment continuously.

6.3.5.3 Virus Hunt

Increase Infection

Kill Virus

In *Virus Hunt* (see Figure 56c), players accumulate points by killing off viruses in the bloodstream. To kill a virus, the user touches it on the screen. However, if she looks at a virus, it infects a new cell, worsening the infection. Therefore, players must kill the viruses using only their peripheral vision. The game was designed to explore two play concepts: the combination of touch input and gaze for gaming and the use of the peripheral vision as a game mechanic. Incorporating touch gestures into gaze interaction creates challenges for eye tracking, because the hands often occlude the user's eyes from the tracker. Whereas this is a problem for critical applications, this challenge can actually be used in game design. For example, games can offer incentives both for touching the screen and ensuring that gaze is tracked, creating a trade-off that the user must balance. The use of peripheral vision is also an exciting opportunity for gaze-enabled games. Whereas in traditional games, players must quickly spot and follow moving targets, our game forces players to go against their instincts and not look at the newly spawned viruses.

6.3.6 Discussion

In this work we augmented the classic arcade machine to create a novel multimodal sensing system. Leveraging this form factor offers several advantages. First, it creates a nostalgic atmosphere that makes the deployment of interactive applications naturally playful. For example, Kruyff et al. also used an arcade machine to increase battery recycling awareness through playful interaction the Alkaline Arcade. Instead of inserting coins, players insert drained batteries, which they then send to recycling [149]. Second, arcade machines provide a very social gaming experience, as they are an object of interest in public environments, often creating a "*honeypot effect*" in which people surround players to cheer and observe [34]. The attention of bystanders can be simultaneously a blessing and a curse. In a survey of the culture around Dance Dance Revolution, Höysniemi identified that whereas this social setting allows them to play with their friends, make new friends and receive support from the audience, it can also lead to bad audience behaviours including disturbing and interfering, making fun, creating a bad atmosphere, etc [118]. Third, arcade machines enable players to use input devices to which they normally would not have access at home. This opens the door to deploying and evaluating multi-modal interaction techniques previously restricted to research labs. The three games we implemented demonstrate how novel game mechanics can stem from a multimodal sensing platform.

We designed Arcade+ to support a variety of combinations of multimodal interactions. In particular, we are interested in gaze-supported interaction techniques for gaming [258,272]. Even though multi-modal gaze-supported techniques have been widely explored in HCI, they are yet to be implemented in gaming contexts. For example, Stellmach et al. and Turner et al. proposed several techniques combining multitouch gestures and gaze [252,271]. Kosunen et al., Cha and Maier and Velloso et al. combined gaze and mid-air gestures [43,146,279]. Our system also allows for combinations of multitouch and mid-air gestures such as the ones explored by Bruder et al. [36,37]. Further, our system supports a variety of foot gestures, such as Simeone et al.'s gestures for 3D interaction, and Velloso et al.'s techniques based on free-form feet movements [242,273]. We hope that by supporting

the sensing of multiple modalities in a single platform, we foster the development of games that incorporate these full-body techniques.

With the advent of personal fabrication, the increasing popularity of the DIY culture and the buzzing of online communities sharing open-source hardware and software, the possibility of end-users building their own custom arcade machines is stronger than ever [52]. Building arcade machines is an active topic in the DIY community [52]. In 1997, Italian programmer Nicola Salmoria released the first version of MAME—Multiple Arcade Machine Emulator—multi-platform emulator that allows users to play old arcade games in modern computers, with the goal of preserving video game heritage. Since then, multiple people have built custom PC cases for MAME machines that mimic old arcade machines [213]. Multiple blogs and websites also provide instructions on how to build DIY arcade machines ranging from Hodges et al.'s pocket-sized .NET Gadgeteer example [115] to Sørensen's detailed guides in his Project MAME's website [246]. All these examples, however, aim at rescuing the original form factor and functionality of old arcade machines. Our goal is to leverage the social and cultural affordances of traditional machines, while extending their original capabilities with modern devices.

The sensing devices in Arcade+ are all commercially available. However, the setup provides many capabilities beyond their conventional ones. Whereas depth cameras such as the Kinect are able to track the user's full body, they require a minimum of 1.4 meters to between the sensor and the player. At this distance, the sensor can only detect coarse gestures, unable to infer detailed information such as finger poses. Further, the eye tracker requires users to be close to the tracker, not only because of the sensor, but also because the further the user is from the display, the smaller the targets are in terms of visual angle. By splitting the sensing of the upper body and lower body, we are able not only to track the whole body close to the display, but also to track precise finger, eye and foot gestures. Another advantage of this setup explored by the game is offloading navigation to the feet in order to free the hands to perform playful and expressive gestures, rather than keeping them on the controls.

6.3.7 Conclusion

In this section, we proposed bringing back arcade machines for the deployment and evaluation of multi-modal games. We designed and built a prototype of Arcade+, a system that offers multitouch gestures, mid-air gestures, foot, and gaze tracking capabilities. We also made the CAD files and instructions available for other researchers to replicate and contribute to the project. To demonstrate the capabilities of our system, we built three games that explore different combinations of input modalities, novel play concepts, and multimodal game mechanics.

We see three distinct directions for future work. First, we will incorporate additional modalities and features into our design. We are working on a second dashboard with another multitouch screen to explore dual-screen games (e.g. *Wii U* games) and indirect touch interaction techniques [241,271]. Another feature that we will explore is how to enable multiplayer sensing. Such setup presents challenges to infra-red-based tracking due to sensor interference, but because of the inherently social nature of arcade machines, it is an interesting direction to pursue. Further, whereas in this iteration, we focused on enabling multimodal sensing, in future work we will also explore novel ways of providing feedback, including stereoscopic 3D and haptics.

The second direction is in developing and evaluating additional games for the platform. The examples we describe in the paper aim at demonstrating the possibilities that our multimodal sensing enables. However, these prototypes only scratch the surface of the

plethora of possible applications. As we explore the affordances of the setup, we will gain a better understanding of how to leverage the characteristics of this setup into game design.

The third direction is deploying and evaluating applications other than games. Whereas our focus is on evaluating games, we see a huge potential for using Arcade+ to investigate other types of multimodal interactive systems. As a research tool, arcade machines bring playfulness and social interaction to system evaluation, while turning prototypes into public objects of interest.

6.4 Conclusion

This chapter was led by the proposition that by treating the body movements as multiple modalities, rather than a single one, we can enable novel user experiences. Due to the inherent difficulty of tracking the whole body up close, we proposed to distribute the sensing across multiple sensors and treat them separately. In our user studies, we evaluated different aspects of the interactions that this sensing setup affords.

First, we conducted an empirical investigation of how the eyes can support mid-air hand gestures in 3D selection. Our study was comprised of three tasks. In the first task, we compared how gaze selection fares against a Virtual Hand and a Raycasting technique. We showed that gaze is significantly faster, and that this advantage lies in the substantially faster target acquisition with the eyes. In the second task, we demonstrated that the eyes are particularly good for tasks in which users must switch between multiple objects in sequence, at the cost of some loss of accuracy. Our third task was designed to demonstrate how the subsequent manipulation affects the selection task. We showed that when users are planning to move the object after selecting they take significantly more time to select it, as they reach out for it, even though the interaction technique does not require them to.

After demonstrating how gaze and mid-air gestures can be combined to provide a superior experience, we set out to explore the ways in which the feet can support three-dimensional interaction tasks. We designed foot-based interaction techniques for the four canonical 3DUI tasks—navigation, object manipulation, selection and system control. Our user study demonstrated that the tasks are easy to use and to learn, but more work is required to fully understand how they can support real tasks.

These two works looked at the upper body and the lower body separately. To bring them all together, we designed and built Arcade+, an augmented arcade machine that enables up close, full-body, multimodal interactions involving mid-air and multitouch hand gestures, feet movements, gaze, and head movements, as well as traditional arcade controls. We demonstrated how the availability of a multimodal sensing platform can inspire novel user experiences through the design on three games.

In regards to our proposition, these works highlight three things. First, the combination of multiple modalities can lead to enhanced task performance. For example, by combining gaze and mid-air gestures, we achieve superior results than with any of the two modalities by themselves. Second, offloading certain tasks to alternative modalities not only frees the hands, but also allows simultaneous control of multiple degrees of freedom. For example, in many 3D modelling applications, the mouse switches between different modes of operation, sometimes controlling the camera, sometimes manipulating the object. By assigning the camera control to the feet, users can simultaneously manipulate the object while changing perspective, similar to how sculptors work. Third, by sensing different modalities separately, we achieve a more detailed and precise full body tracking. For example, the Kinect is able to track the whole body at a distance. However, by focusing each sensor in one specific body part, we achieve not only a higher granularity of detail (e.g. finger movement and hand gestures), but also enable interactions closer to the display.

7 CONCLUSION

"Research is to see what everybody else has seen, and to think what nobody else has thought."

Albert Szent-Gyorgyi

In this thesis, we investigated the research space of Bodily Interaction in three directions: implicit interaction, lower limb interaction and multimodal interaction. Our work was guided by the propositions that there is more to be inferred by natural users' movements and postures; that the lower limbs can provide an effective means of interacting with computers beyond assistive technology; and that by treating body movements as multiple modalities, rather than a single one, we can enable novel user experiences. This chapter summarises our contributions, discusses the lessons learned from our investigations, and suggests directions for future work.

7.1 Reflections on the Research Questions

Our propositions led to the design of three research questions that guided the direction of this thesis. The first question was **how do we specify movements for quality analysis?** Based on our discussions about quality, we reached the conclusion that when analysing quality, it is of utmost importance to first establish a baseline against which we can compare further executions of that movement. In the weight lifting domain, this seemed rather straightforward at first. Exercises are well defined and described, and performing the movement accordingly is often seen as indicative of expertise in the area. However, what we found is that even for common exercises there is a huge variation in the movement execution not only between different experts but between how experts say they perform the movement and how they actually do. This led us through an iterative exploration of different approaches for achieving a specification that not only could be used for subsequent analyses of executions by other people, but that supported the tacit nature of movement communication. In the second domain we investigated, Affective Computing, how emotional states are conveyed through movement is an even more ambiguous specification problem. On one hand we, as humans, are instinctively able to recognise and distinguish different

affective states of other people by how their bodies behave. On the other hand, explicitly and unambiguously specifying which movements and postures are connected to these different states is a challenging problem and an open question even in the Affective Sciences. Our approach was therefore to break whole body movements and postures into their finer-grained components. Similarly to the approach taken to analyse weight lifting exercises in Chapter 2, our approach for labelling body expressions in Chapter 3 relied on users demonstrating different small, canonical movements that together create affective expressions.

The second question was **how can the feet be used for direct input to interactive systems?** This question was guided by the insight that even though foot-based interfaces exist since the inception of Human-Computer Interaction, there was no comprehensive understanding of how the feet work as an input modality. Again, our approach was to look not at the body as a whole but at an individual body part. However, whereas in Chapters 2 and 3 we analysed different joints and bones independently of each other, but still observed the whole body; in Chapters 4 and 5, we specifically focused on the lower limbs. Our investigations led to several new insights, including that even though feet movements are considerably slower and less accurate than the hands, they still perform similarly to other input devices less used than the mouse, such as game controllers and touchpads. Other findings include that both feet perform similarly and that the direction of movement impacts the pointing performance.

At this stage, having looked at the body as whole in Chapters 2 and 3, and specifically at the feet in Chapters 4 and 5, our next goal was to understand how different body parts, understood as independent input modalities can collaborate to complete certain tasks. Therefore, our third question was **how can combinations of input modalities support complex interactive tasks?** We first explored gaze and mid-air gestures, modalities that had been investigated extensively by themselves, but not together, in the domain of 3D user interfaces. We showed how their inherently different characteristics can overcome their individual shortcomings, leading to a superior user experience. We then applied our insights from Chapters 4 and 5 into designing interaction techniques for the same domain. Finally, we brought together these two sensing setups and their corresponding interaction techniques into a full body sensing system for gaming in the form of an arcade machine.

The explorations of the research questions in the thesis highlight an approach of investigating body movements that rather than looking at full body interaction as a black box, explores the design opportunities and challenges of each individual body part as an input modality and how they can together enable novel user experiences.

7.2 Contributions

Each direction that we explored yielded distinct contributions, summarised in the following:

1. **Implicit Interaction:** We contribute a formalisation of the concepts of quality and of qualitative activity recognition, as well as systems and algorithms that implement different approaches to them. Based on the literature on quality in other areas, we define it as *the adherence of the execution of an activity to its specification*, and we define a qualitative activity recognition system as *a system that observes the user's execution of an activity and compares it to a specification*. As these definitions highlight the need for an activity specification, we contribute three approaches for extracting it: (1) demonstrating the correct execution, as well as possible mistakes; (2) manually specifying a model that analyses incoming data in an object-oriented

fashion; (3) demonstrating the correct execution and extracting the activity model automatically. All approaches successfully encoded movement information and provided feedback on the quality of execution. We show that our third approach combines the advantages of allowing users to encode tacit information by demonstrating the movement without having to estimate movement parameters explicitly (we found that experts' estimates of such parameters are significantly different to their performance), while still allowing users to manually modify movement parameters if they wish.

2. **Lower Body Interaction:** We contribute a comprehensive survey of lower body human-computer interaction and an empirical characterisation of unconstrained foot movements for computing input. We described the human factors of lower limb interaction; systems that involve the lower body; and the different types of interactions that the feet enable. Our survey also sets the scene for an in-depth analysis of some of the fundamental aspects of foot-based interaction. To enable computing input with free-form foot movements in a desktop setting, we built a foot tracker based on a Kinect sensor mounted under the desk. With this equipment we conducted a series of studies that evaluate different aspects of such interaction. First, we built Fitts's Law performance models for unconstrained foot pointing. We then showed that left and right movements are easier and faster than forward and backward ones. We also showed that the footedness of users does not significantly impact the performance. Our results indicate that the mapping between foot controlled parameters and visual feedback strongly influences the interaction, and that the overhead incurred by the addition of a foot controlled parameter on hand-based interaction is not significant.
3. **Multimodal Interaction:** We contribute an exploration of the novel user experiences that are afforded by a multi-sensor full body tracking platform. We demonstrated that the eyes can effectively speed up other mid-air gesture-based 3D selection techniques and that the feet can provide an alternative for the canonical 3D interaction tasks (navigation, object manipulation, selection and system control). Finally, to enable playful multimodal full-body up close interactions, we built an augmented arcade machine. Our prototype is able to simultaneously sense mid-air gestures, multitouch gestures, gaze position, head movements, feet movements, as well as to receive input from traditional arcade controls.

7.3 Lessons Learned

So far, we have presented the individual results from our research. We now reflect on the lessons learned in these years of research on bodily interaction and how they relate to the propositions we started with.

- **Movement information is often communicated implicitly, and systems that support this communication process should take this into account.** A considerable amount of information is transmitted tacitly through our body movements and postures. In our investigations on supporting the communication process between expert weight lifters and novices, we found the best results when our system emulated this process. This is a powerful insight in the design of teaching systems in general. By replicating a real world communication process we were able to not only provide a more satisfying user experience, but also contribute an innovative algorithm to model and analyse movement information. These insights from Chapter 2 were also incorporated in the design of the system presented in

Chapter 3, as we teach the system to label body expressions by demonstrating them beforehand.

- **Quality analysis brings another layer of information to activity recognition**. Despite a considerable amount of work having been conducted in activity recognition, little attention has been given to quality. However, quality recognition and activity recognition can strongly benefit each other. On one hand, recognising which activity is being performed can be a first step to recognising the activity's quality. For example, in a gym environment, an activity classifier can determine which exercise the user is doing and trigger the appropriate quality analyser. On the other hand, quality data can be used to infer further contextual cues. For example, if the system detects a poor quality of activity, it can infer that the user is tired or not concentrated enough. Chapters 2 and 3 both investigate movement quality under our definition, but from different angles. Whereas in Chapter 2, the movement quality relates to how well an exercise is being performed, in Chapter 3, we focus on the emotional quality of the movement. Considering our definition of quality as how close the movement is to a certain specification, we can model the affective content of a movement of where it is in a large possibility space of combinations of expressions.
- **Analysing affective body expressions at a higher level can be advantageous across multiple research areas**. Even if previous research has shown that it is possible to classify emotions from body movements with data-driven classifiers, this is of little use for researchers outside affective computing. By working with behaviour labels, such as the ones we extracted with AutoBAP, research in affective computing can lead to insights that can benefit psychology and emotion research, and vice versa. A standard set of labels not only enables a cross-disciplinary common language, but also enables researchers to explain the reasons behind the classification of a certain expression. In a data-driven classifier, this kind of information is encoded into the classifier, but not in an explicit way.
- **The lower body can be used as a valuable complementary modality for computing input**. As we have shown in our survey, many works have employed the legs and feet for HCI, but with different backgrounds and purposes. In our survey, we achieved several insights, including that the feet excel at performing simple tasks, feet interfaces can assist the hands rather than replace them, the performance of the feet might not be as bad as people think, that foot-based input lends itself well to wearable computing, that we still lack the understanding of how these interfaces work in the real world, and that interactive systems can further benefit from what the feet tell about users' psychological states. We hope that with this increased understanding of the possible roles and capabilities, we see more and more interesting applications for feet, both as a primary and as a complementary modality for computing input.
- **Treating different body parts as different modalities can inspire many novel multimodal applications.** Before starting this work, we usually thought of body movements as a single modality. As we began to conceptually separate them, we started to see how they can be used in cleverer ways to enhance existing interaction techniques (e.g. how gaze can speed up mid-air selection), create novel techniques (e.g. our foot-based 3DUI techniques), and inspire exciting new applications (e.g. our multimodal games). Throughout the thesis we took different approaches to deconstructing body movements: in Chapters 2 and 3, we created models for individual joints and segments; in Chapters 4 and 5, we investigated the feet as an individual body movement. However, in Chapter 6, we started bringing these insights together and in a sense "reconstructing" the whole body, but now as a sum

of parts rather than a black box. Our arcade demonstrates this concept, but enabling full body interaction, but treating each body part as an independent modality.

7.4 Future Directions

This thesis opened the doors to several new research directions. In this section, we suggest a few of these as potential future work.

The work on implicit interaction stemmed from an interest in Ubiquitous Computing and the vision of an interconnected web of devices that collaborate to create novel user experiences. Given this goal, the gym offers several advantages as a testbed context. It is a closed environment, where users perform a finite set of tasks, using a determined set of objects and equipment, and they can strongly benefit from receiving feedback on their activities. However, despite all these advantages, very little work had been conducted in adding computational support to weight lifting.

Our vision was that of a gym where every dumbbell, barbell and weight was augmented with microcontrollers, sensors, actuators and networking capabilities. These devices could reason amongst themselves and the user's exercise program to provide feedback on the activities and suggest other exercises. Whereas we proposed several solutions for the problem of exercise analysis and feedback, there are many other areas that must be investigated to achieve this vision.

First, future work should make an effort in deploying such systems out in the wild. Despite our user studies demonstrating that the systems work in a controlled setting, the gym environment presents much more complex challenges, including social (e.g. turn-taking in using the equipment), technical (e.g. making the systems robust enough to survive high impact with the floor) and ethical (e.g. if the system leads to a user injuring himself). Second, we tried to support the communication of movement information when experts and novices were far apart, but there are many opportunities for designing systems that assist this same communication process when they are co-located, including novel on-body visualisations of important movement aspects, and motivational technologies, such as gamification, to improve performance. Third, we explored the domain of weight lifting, but these techniques could possibly be extended into other domains. Since the publication of our work, other researchers have expanded these areas. For example, Spina et al. used an IMU approach that also allowed a trainer to demonstrate the exercise for the user to perform, but for physical rehabilitation exercises [248]. Tang et al. built *Physio@Home*, a system for the same domain using an augmented mirror [260]. With a similar aim of sharing movement information, Anderson et al. built *YouMove*, a system that also provides feedback through an augmented mirror [3]. Additional domains to be explored include dance, other sports, and games.

Chapter 3 contributed a system to support the recognition of emotions from affective body expressions, by working on a higher level. However, we stopped at the labelling of objective nonverbal cues, without making any judgement on their emotional content. We see three general directions for future work: towards the coding system itself, towards the automatic annotation and towards affect recognition. In terms of the coding system, future work should look into expanding the labels for the lower body. As shown in our survey of foot-based interactions, the lower limbs can offer significant insights into our psychological states, but the current manual for the Body Action and Posture coding system is more focused on the upper body, with few labels to classify lower limb behaviours. In terms of the automatic annotation, more work needs to be conducted into making the movement capture less obtrusive. The motion capture suit we used is highly accurate, but requires heavy user

instrumentation, which can affect their natural behaviours. Moreover, it does not cover certain body parts, such as the fingers, so other solutions for tracking smaller movements are necessary. Finally, as the coding system matures, research from Psychology will provide more insights into how the individual bodily nonverbal cues relate to particular emotional states, making their automatic annotation a valuable tool in recognising affect automatically.

Our work on foot interaction revealed several open research questions. We evaluated certain aspects of foot interaction, but we only focused on abstract tasks. More work must be conducted on concrete applications. Also, a longer term study is necessary to understand how users' performances change over time. All our participants were novices to foot interaction, so it is still unclear if our results generalise to expert users. Finally, we believe that there are several opportunities for further work on using the feet to support the hands. We hope with the development of our multimodal arcade machine, we will see more interaction techniques that combine the feet with other input modalities.

We built Arcade+ to demonstrate how a multimodal sensing platform can inspire novel user experiences in the form of new game mechanics. However, the possibilities for future use of this platform are not limited by games. We envision using it as a platform to deploy and evaluate other interactive systems in public spaces. Another interesting aspect to investigate is the social impact of this form factor. Arcade machines suggest a playful, social and casual nature, which can positively influence users' perceptions and experiences.

Body movements allow us to experience digital content in a similar way to how we experience the physical world around us. Our research has tackled many aspects of this kind of interaction. We showed that observing users' implicit movements, incorporating the lower body in interactive systems and combining input from multiple body parts can create novel and exciting user experiences and we hope that future commercial systems can also take these aspects into account.

8 REFERENCES

1. Abowd, G.D., Dey, A.K., Brown, P.J., Davies, N., Smith, M., and Steggles, P. Towards a Better Understanding of Context and Context-Awareness. *Proceedings of the 1st International Symposium on Handheld and Ubiquitous Computing*, Springer-Verlag (1999), 304–307.

2. Alexander, J., Han, T., Judd, W., Irani, P., and Subramanian, S. Putting your best foot forward: investigating real-world mappings for foot-based gestures. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM (2012), 1229–1238.

3. Anderson, F., Grossman, T., Matejka, J., and Fitzmaurice, G. YouMove: enhancing movement training with an augmented reality mirror. *Proceedings of the 26th annual ACM symposium on User interface software and technology*, ACM (2013), 311–320.

4. Arbib, M.A. Perceptual structures and distributed motor control. *Comprehensive Physiology*, (1981).

5. Argelaguet, F. and Andujar, C. A survey of 3D object selection techniques for virtual environments. *Computers & Graphics 37*, 3 (2013), 121–136.

6. Argyle, M. *Bodily communication*. Routledge, 1988.

7. Augsten, T., Kaefer, K., Meusel, R., et al. Multitoe: high-precision interaction with back-projected floors based on high-resolution multi-touch input. *Proceedings of the 23nd annual ACM symposium on User interface software and technology*, ACM (2010), 209–218.

8. Baca, A. and Kornfeind, P. Rapid feedback systems for elite sports training. *Pervasive Computing, IEEE 5*, 4 (2006), 70–76.

9. Bächlin, M., Förster, K., and Tröster, G. SwimMaster: a wearable assistant for swimmer. *Proceedings of the 11th international conference on Ubiquitous computing*, ACM (2009), 215–224.

10. Balakrishnan, R., Fitzmaurice, G., Kurtenbach, G., Singh, K., and East, K.S. Exploring interactive curve and surface manipulation using a bend and twist sensitive input strip. *Proc. I3D '99*, ACM (1999), 111–118.

11. Bannach, D., Lukowicz, P., and Amft, O. Rapid prototyping of activity recognition applications. *Pervasive Computing, IEEE 7*, 2 (2008), 22–31.

12. Bänziger, T. and Scherer, K.R. Introducing the geneva multimodal emotion portrayal (gemep) corpus. *Blueprint for affective computing: A sourcebook*, (2010), 271–294.

13. Barnes, R.M., Hardaway, H., and Podolsky, O. Which pedal is best. *Factory Management and Maintenance 100*, (1942), 98–99.

14. Barnett, R.L. Foot controls: Riding the pedal. *The Ergonomics Open Journal 2*, (2009), 13–16.

15. Barrett, L.F. Variety is the spice of life: A psychological construction approach to understanding variability in emotion. *Cognition and Emotion 23*, 7 (2009), 1284–1306.

16. Bazzo, J.J. and Lamar, M.V. Recognizing facial actions using Gabor wavelets with neutral face average difference. *Automatic Face and Gesture Recognition, 2004. Proc.. Sixth IEEE International Conference on*, IEEE (2004), 505–510.

17. Beetz, M., Kirchlechner, B., and Lames, M. Computerized real-time analysis of football games. *Pervasive Computing, IEEE 4*, 3 (2005), 33–39.

18. Benko, H. Beyond flat surface computing: challenges of depth-aware and curved interfaces. *Proceedings of the 17th ACM international conference on Multimedia*, ACM (2009), 935–944.

19. Bianchi-Berthouze, N., Cairns, P., Cox, A., Jennett, C., and Kim, W.W. On posture as a modality for expressing and recognizing emotions. *Emotion and HCI workshop at BCS HCI London*, (2006).

20. Bianchi-Berthouze, N., Kim, W.W., and Patel, D. Does body movement engage you more in digital game play? And Why? In *Affective Computing and Intelligent Interaction*. Springer, 2007, 102–113.

21. Bianchi-Berthouze, N. Understanding the role of body movement in player engagement. *Human–Computer Interaction 28*, 1 (2013), 40–75.

22. Bieber, G. and Diener, H. Stepman-a new kind of music interaction. Proceedings of the International Conference on Human-Computer Interaction (HCI International), (2005).

23. Billinghurst, M. and Buxton, B. Gesture based interaction. *Human Input to Computer Systems: Theories, Techniques and Technology 24*, (2011).

24. Birdwhistell, R.L. *Kinesics and context: Essays on body motion communication*. University of Pennsylvania press, 1970.

25. Blair, S.N. Physical inactivity: the biggest public health problem of the 21st century. *British journal of sports medicine 43*, 1 (2009), 1–2.

26. Borazio, M. and Van Laerhoven, K. Improving Activity Recognition Without Sensor Data: A Comparison Study of Time Use Surveys. *Proceedings of the 4th Augmented Human International Conference*, ACM (2013), 108–115.

27. Bowman, D.A. and Hodges, L.F. An evaluation of techniques for grabbing and manipulating remote objects in immersive virtual environments. *Proceedings of the 1997 symposium on Interactive 3D graphics*, ACM (1997), 35–ff.

28. Bowman, D.A., Kruijff, E., LaViola Jr, J.J., and Poupyrev, I. An introduction to 3-D user interface design. *Presence-Teleop. Virt. 10*, 1 (2001), 96–108.

29. Bowman, D.A., Kruijff, E., LaViola Jr, J.J., and Poupyrev, I. *3D user interfaces: theory and practice*. Addison-Wesley, 2004.

30. Bowman, D.A., McMahan, R.P., and Ragan, E.D. Questioning Naturalism in 3D User Interfaces. *Communications of the ACM 55*, 9 (2012), 78–88.

31. Bränzel, A., Holz, C., Hoffmann, D., et al. GravitySpace: tracking users and their poses in a smart room using a pressure-sensing floor. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM Press (2013), 725–734.

32. Breiman, L. Bagging predictors. *Machine learning 24*, 2 (1996), 123–140.

33. Breiman, L. Random forests. *Machine learning 45*, 1 (2001), 5–32.

34. Brignull, H. and Rogers, Y. Enticing people to interact with large public displays in public spaces. *Proceedings of INTERACT*, (2003), 17–24.

35. Brogan, D.C., Metoyer, R.A., and Hodgins, J.K. Dynamically simulated characters in virtual environments. *Computer Graphics and Applications, IEEE 18*, 5 (1998), 58–69.

36. Bruder, G., Steinicke, F., and Stuerzlinger, W. Touching the void revisited: Analyses of touch behavior on and above tabletop surfaces. In *Human-Computer Interaction–INTERACT 2013*. Springer, 2013, 278–296.

37. Bruder, G., Steinicke, F., and Sturzlinger, W. To touch or not to touch?: comparing 2D touch and 3D mid-air interaction on stereoscopic tabletop surfaces. *Proc. SUI '13*, ACM (2013), 9–16.

38. Bruno, L., Pereira, J., and Jorge, J. A new approach to walking in place. In *Human-Computer Interaction–INTERACT 2013*. Springer, 2013, 370–387.

39. Bulling, A., Blanke, U., and Schiele, B. A tutorial on human activity recognition using body-worn inertial sensors. *ACM Computing Surveys (CSUR) 46*, 3 (2014), 33.

40. Buttussi, F. and Chittaro, L. MOPET: A context-aware and user-adaptive wearable system for fitness training. *Artificial Intelligence in Medicine 42*, 2 (2008), 153–163.

41. Calinon, S. Robot programming by demonstration. In *Springer handbook of robotics*. Springer, 2008, 1371–1394.

42. Carrozza, M.C., Persichetti, A., Laschi, C., et al. A Wearable Biomechatronic Interface for Controlling Robots with Voluntary Foot Movements. *Mechatronics, IEEE/ASME Transactions on 12*, 1 (2007), 1–11.

43. Cha, T. and Maier, S. Eye gaze assisted human-computer interaction in a hand gesture controlled multi-display environment. *Proceedings of the 4th Workshop on Eye Gaze in Intelligent Human Machine Interaction*, ACM (2012), 13.

44. Chan, A.O.K., Chan, A.H.S., Ng, A.W.Y., and Luk, B.L. A Preliminary Analysis of Movement Times and Subjective Evaluations for a Visually-Controlled Foot-Tapping Task on Touch Pad Device. *Proceedings of the International MultiConference of Engineers and Computer Scientists*, (2010), 17–19.

45. Chan, K.W.L.L. and Chan, A.H.S.S. Spatial stimulus–response (SR) compatibility for foot controls with visual displays. *International Journal of Industrial Ergonomics 39*, 2 (2009), 396–402.

46. Chang, K., Chen, M.Y., and Canny, J. Tracking Free-Weight Exercises. *UbiComp 2007 Ubiquitous Computing 4717*, (2007), 19–37.

47. Chang, Y.-J., Chen, S.-F., and Huang, J.-D. A Kinect-based system for physical rehabilitation: A pilot study for young adults with motor disabilities. *Research in developmental disabilities 32*, 6 (2011), 2566–2570.

48. Chartrand, T.L. and Bargh, J.A. The chameleon effect: The perception–behavior link and social interaction. *Journal of personality and social psychology 76*, 6 (1999), 893.

49. Chen, M., Huang, B., and Xu, Y. Intelligent shoes for abnormal gait detection. *Robotics and Automation, 2008. ICRA 2008. IEEE International Conference on*, IEEE (2008), 2019–2024.

50. Choi, I. and Ricci, C. Foot-mounted gesture detection and its application in virtual environments. *Proc. SCM '97*, IEEE (1997), 4248–4253.

51. Chuang, C.F. and Shih, F.Y. Recognizing facial action units using independent component analysis and support vector machine. *Pattern recognition 39*, 9 (2006), 1795–1798.

52. Clair, J.S. Project Arcade: Build Your Own Arcade Machine. John Wiley and Sons, 2011.

53. Cohen, J. and others. A coefficient of agreement for nominal scales. *Educational and psychological measurement 20*, 1 (1960), 37–46.

54. Cournia, N., Smith, J.D., and Duchowski, A.T. Gaze-vs. hand-based pointing in virtual environments. *CHI'03 extended abstracts on Human factors in computing systems*, ACM (2003), 772–773.

55. Crosby, P.B. Quality is free: The art of making quality certain. McGraw-Hill New York, 1979.

56. Crossan, A., Brewster, S., and Ng, A. Foot tapping for mobile interaction. *Proceedings of the 24th BCS Interaction Specialist Group Conference*, British Computer Society (2010), 418–422.

57. Cypher, A. and Halbert, D.C. *Watch what I do: programming by demonstration*. MIT press, 1993.

58. Cypher, A. and Smith, D.C. KidSim: end user programming of simulations. *Proceedings of the SIGCHI conference on Human factors in computing systems*, ACM Press/Addison-Wesley Publishing Co. (1995), 27–34.

59. D'Mello, S. and Calvo, R.A. Beyond the Basic Emotions: What Should Affective Computing Compute? *CHI '13 Extended Abstracts on Human Factors in Computing Systems*, ACM (2013), 2287–2294.

60. Dael, N., Mortillaro, M., and Scherer, K. The Body Action and Posture Coding System (BAP): Development and Reliability. *Journal of Nonverbal Behavior 36*, 2 (2012), 97–121.

61. Dael, N., Mortillaro, M., and Scherer, K.R. The Body Action and Posture coding system (BAP): Development and reliability. *Journal of Nonverbal Behavior*, (2012), 1–25.

62. Daiber, F., Schöning, J., and Krüger, A. Whole body interaction with geospatial data. *Proc. SG '09*, Springer (2009), 81–92.

63. Dargent-Paré, C., De Agostini, M., Mesbah, M., and Dellatolas, G. Foot and eye preferences in adults: Relationship with handedness, sex and age. *Cortex 28*, 3 (1992), 343–351.

64. Darken, R.P., Cockayne, W.R., and Carmein, D. The omni-directional treadmill: a locomotion device for virtual worlds. *Proceedings of the 10th annual ACM symposium on User interface software and technology*, ACM (1997), 213–221.

65. Dawe, E.J.C. and Davis, J. (vi) Anatomy and biomechanics of the foot and ankle. *Orthopaedics and Trauma 25*, 4 (2011), 279–286.

66. Dearman, D., Karlson, A., Meyers, B., and Bederson, B. Multi-modal text entry and selection on a mobile device. *Proceedings of Graphics Interface 2010*, Canadian Information Processing Society (2010), 19–26.

67. Dempster, W.T. Space requirements of the seated operator: geometrical, kinematic, and mechanical aspects of the body, with special reference to the limbs. (1955).

68. Dey, A.K., Hamid, R., Beckmann, C., Li, I., and Hsu, D. a CAPpella: programming by demonstration of context-aware applications. *Proceedings of the SIGCHI conference on Human factors in computing systems*, ACM (2004), 33–40.

69. Donovan, T. and Garriott, R. *Replay: The history of video games*. Yellow Ant Lewes, UK, 2010.

70. Drossis, G., Grammenos, D., Bouhli, M., Adami, I., and Stephanidis, C. Comparative Evaluation among Diverse Interaction Techniques in Three Dimensional Environments. In *Distributed, Ambient, and Pervasive Interactions*. Springer, 2013, 3–12.

71. Drury, C.G. Application of Fitts' Law to foot-pedal design. *Human Factors: The Journal of the Human Factors and Ergonomics Society 17*, 4 (1975), 368–373.

72. Dutta, T. Evaluation of the Kinect sensor for 3-D kinematic measurement in the workplace. *Applied ergonomics 43*, 4 (2012), 645–649.

73. Ekman, P. and Friesen, W. Facial Action Coding System: A Technique for the Measurement of Facial Movement. Consulting Psychologists Press, Palo Alto, 1978.

74. Ekman, P., Friesen, W.V., and Hager, J.C. *Facial action coding system*. A Human Face Salt Lake City, 2002.

75. Ekman, P. and Friesen, W.V. Constants across cultures in the face and emotion. *Journal of personality and social psychology 17*, 2 (1971), 124.

76. Ekman, P. and Friesen, W.V. Facial action coding system. (1977).

77. Ekman, P. An argument for basic emotions. *Cognition & emotion 6*, 3-4 (1992), 169–200.

78. Engelbart, D. Doug Engelbart Discusses Mouse Alternatives. 1984. ftp://ftp.cs.utk.edu/pub/shuford/terminal/engelbart_mouse_alternatives.html.

79. English, W.K., Engelbart, D.C., and Berman, M.L. Display-selection techniques for text manipulation. *Human Factors in Electronics, IEEE Transactions on*, 1 (1967), 5–15.

80. Ermes, M., Parkka, J., Mantyjarvi, J., and Korhonen, I. Detection of daily activities and sports with wearable sensors in controlled and uncontrolled conditions. *Information Technology in Biomedicine, IEEE Transactions on 12*, 1 (2008), 20–26.

81. Feasel, J., Whitton, M.C., and Wendt, J.D. LLCM-WIP: Low-latency, continuous-motion walking-in-place. *3D User Interfaces, 2008. 3DUI 2008. IEEE Symposium on*, IEEE (2008), 97–104.

82. Fischer, C., Talkad Sukumar, P., and Hazas, M. Tutorial: implementation of a pedestrian tracker using foot-mounted inertial sensors. *IEEE Pervasive Computing 12*, 2 (2013), 17–27.

83. Fitts, P.M. The information capacity of the human motor system in controlling the amplitude of movement. *Journal of experimental psychology 47*, 6 (1954), 381.

84. Fleiss, J.L., Levin, B., and Paik, M.C. The measurement of interrater agreement. *Statistical methods for rates and proportions 2*, (1981), 212–236.

85. Ford, M. Antique Arcade Games-Mike Munves Catalog 1939-1962. Classic Arcade Grafix, 2009.

86. Frey, M. CabBoots: shoes with integrated guidance system. *Proceedings of the 1st international conference on Tangible and embedded interaction*, ACM (2007), 245–246.

87. Gallagher, M. Pain In The Mass: Ten Most Common Causes of Training Injury. *Muscle & Fitness 57*, (1996), 68.

88. Garcia, F.P. and Vu, K.-P.L. Effects of Practice with Foot-and Hand-Operated Secondary Input Devices on Performance of a Word-Processing Task. In *Human Interface and the Management of Information. Designing Information Environments*. Springer, 2009, 505–514.

89. Garcia, F.P. and Vu, K.-P.L. Effectiveness of hand- and foot-operated secondary input devices for word-processing tasks before and after training. *Computers in Human Behavior 27*, 1 (2011), 285–295.

90. Gelder, B. de. Why bodies? Twelve reasons for including bodily expressions in affective neuroscience. *Philosophical Transactions of the Royal Society B: Biological Sciences 364*, 1535 (2009), 3475–3484.

91. Göbel, F., Klamka, K., Siegel, A., Vogt, S., Stellmach, S., and Dachselt, R. Gaze-supported foot interaction in zoomable information spaces. *CHI '13 Extended Abstracts on Human Factors in Computing Systems on - CHI EA '13*, ACM Press (2013), 3059.

92. Göbel, F., Klamka, K., Siegel, A., Vogt, S., Stellmach, S., and Dachselt, R. Gaze-supported foot interaction in zoomable information spaces. *CHI'13 Extended Abstracts on Human Factors in Computing Systems*, ACM (2013), 3059–3062.

93. Gregersen, T.S. Nonverbal cues: Clues to the detection of foreign language anxiety. *Foreign Language Annals 38*, 3 (2005), 388–400.

94. Griffin, H.J., Aung, M.S., Romera-Paredes, B., et al. Laughter type recognition from whole body motion. *Affective Computing and Intelligent Interaction (ACII), 2013 Humaine Association Conference on*, IEEE (2013), 349–355.

95. Griffith, N. and Fernström, M. LiteFoot: A floor space for recording dance and controlling media. *Proceedings of the 1998 International Computer Music Conference*, (1998), 475–481.

96. Grønbæk, K., Iversen, O.S., Kortbek, K.J., Nielsen, K.R., and Aagaard, L. IGameFloor: a platform for co-located collaborative games. *Proceedings of the international conference on Advances in computer entertainment technology*, ACM (2007), 64–71.

97. Gunes, H. and Pantic, M. Dimensional emotion prediction from spontaneous head gestures for interaction with sensitive artificial listeners. *Intelligent virtual agents*, Springer (2010), 371–377.

98. Halbert, D.C. Programming by example. 1984.

99. Hales, J., Rozado, D., and Mardanbegi, D. Interacting with Objects in the Environment by Gaze and Hand Gestures. *ECEM*, (2011).

100. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I.H. The WEKA data mining software: an update. *SIGKDD Explor. Newsl. 11*, 1 (2009), 10–18.

101. Hall, M.A. Correlation-based feature selection for machine learning. 1999.

102. Hamm, J., Kohler, C.G., Gur, R.C., and Verma, R. Automated Facial Action Coding System for dynamic analysis of facial expressions in neuropsychiatric disorders. *Journal of neuroscience methods 200*, 2 (2011), 237–256.

103. Hammerla, N.Y., Plötz, T., Andras, P., and Olivier, P. Assessing motor performance with pca. *Proceedings of the International Workshop on Frontiers in Activity Recognition using Pervasive Sensing*, (2011), 18–23.

104. Han, T., Alexander, J., Karnik, A., Irani, P., and Subramanian, S. Kick: investigating the use of kick gestures for mobile interactions. *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services*, ACM Press (2011), 29–32.

105. Harrigan, J.A., Oxman, T.E., and Rosenthal, R. Rapport expressed through nonverbal behavior. *Journal of Nonverbal Behavior 9*, 2 (1985), 95–110.

106. Havens, T.C., Alexander, G.L., Abbott, C., Keller, J.M., Skubic, M., and Rantz, M. Contour tracking of human exercises. *Computational Intelligence for Visual Intelligence, 2009. CIVI'09. IEEE Workshop on*, IEEE (2009), 22–28.

107. Herndon, K.P., Dam, A. van, and Gleicher, M. The challenges of 3D interaction: a CHI '94 workshop. *SIGCHI Bull. 26*, 4 (1994), 36–43.

108. Hey, J. and Carter, S. Pervasive computing in sports training. *Pervasive Computing, IEEE 4*, 3 (2005), 54.

109. Hick, W.E. On the rate of gain of information. *Quarterly Journal of Experimental Psychology 4*, 1 (1952), 11–26.

110. Higuchi, H. and Nojima, T. Shoe-shaped i/o interface. Adjunct proceedings of the 23nd annual ACM symposium on User interface software and technology, ACM (2010), 423–424.

111. Hinckley, K., Jacob, R., and Ware, C. Input/output devices and interaction techniques. In A.B. Tucker, ed., *CRC Computer Science and Engineering Handbook*. CRC Press LLC: Boca Raton, FL. to appear, 2004, 20.1–20.32.

112. Hinckley, K., Pausch, R., Goble, J.C., and Kassell, N.F. A survey of design issues in spatial input. *Proceedings of the 7th annual ACM symposium on User interface software and technology*, ACM (1994), 213–222.

113. Hinckley, K. and Wigdor, D. Input technologies and techniques. The human-computer interaction handbook: fundamentals, evolving technologies and emerging applications, (2002), 151–168.

114. Hockman, J.A., Wanderley, M.M., and Fujinaga, I. Real-time phase vocoder manipulation by runner's pace. *Proc. Int. Conf. on New Interfaces for Musical Expression (NIME)*, (2009).

115. Hodges, S., Taylor, S., Villar, N., Scott, J., and Helmes, J. Exploring physical prototyping techniques for functional devices using. NET gadgeteer. *Proceedings of the 7th*

International Conference on Tangible, Embedded and Embodied Interaction, ACM (2013), 271–274.

116. Hoffmann, E.R. A comparison of hand and foot movement times. *Ergonomics 34*, 4 (1991), 397–406.

117. Hollerbach, J.M. Locomotion interfaces. Handbook of virtual environments: Design, implementation, and applications, (2002), 239–254.

118. Hoysniemi, J. International Survey on the Dance Dance Revolution Game. *Comput. Entertain. 4*, 2 (2006).

119. Hu, R.Z.-L., Hartfiel, A., Tung, J., Fakih, A., Hoey, J., and Poupart, P. 3D Pose tracking of walker users' lower limb with a structured-light camera on a moving platform. *Computer Vision and Pattern Recognition Workshops (CVPRW), 2011 IEEE Computer Society Conference on*, IEEE (2011), 29–36.

120. Huang, B., Chen, M., Ye, W., and Xu, Y. Intelligent Shoes for Human Identification. *Robotics and Biomimetics, 2006. ROBIO '06. IEEE International Conference on*, (2006), 601–606.

121. Huang, W. and Alem, L. Supporting hand gestures in mobile remote collaboration: a usability evaluation. *Proceedings of the 25th BCS Conference on Human-Computer Interaction*, British Computer Society (2011), 211–216.

122. Huber, B. Foot position as indicator of spatial interest at public displays. *CHI '13 Extended Abstracts on Human Factors in Computing Systems on - CHI EA '13*, ACM Press (2013), 2695.

123. Hudson, S.E., Harrison, C., Harrison, B.L., and LaMarca, A. Whack gestures: inexact and inattentive interaction with mobile devices. *Proceedings of the fourth international conference on Tangible, embedded, and embodied interaction*, ACM (2010), 109–112.

124. Hyman, R. Stimulus information as a determinant of reaction time. *Journal of experimental psychology 45*, 3 (1953), 188.

125. Iskandar, P., Hanum, Y., Gilbert, L., and Wills, G. The design of effective feedback in computer-based sport training. *Proceedings of the 7th International Symposium on Computer Science in Sport*, (2009), 1–13.

126. Iwata, H., Yano, H., and Nakaizumi, F. Gait master: A versatile locomotion interface for uneven virtual terrain. *Virtual Reality, 2001. Proceedings. IEEE*, IEEE (2001), 131–137.

127. Jacob, R.J. What you look at is what you get: eye movement-based interaction techniques. *Proceedings of the SIGCHI conference on Human factors in computing systems*, ACM (1990), 11–18.

128. Jalaliniya, S., Smith, J., Sousa, M., Büthe, L., and Pederson, T. Touch-less interaction with medical images using hand & foot gestures. *Proceedings of the 2013 ACM conference on Pervasive and ubiquitous computing adjunct publication - UbiComp '13 Adjunct*, ACM Press (2013), 1265.

129. Jean, F. and Albu, A.B. The visual keyboard: Real-time feet tracking for the control of musical meta-instruments. *Signal Processing: Image Communication 23*, 7 (2008), 505–515.

130. Jeong, I.-W., Seo, Y.-H., and Yang, H.S. Effective humanoid motion generation based on programming-by-demonstration method for entertainment robotics. *Virtual Systems and Multimedia (VSMM), 2010 16th International Conference on*, IEEE (2010), 289–292.

131. Jota, R., Lopes, P., Wigdor, D., and Jorge, J.A. Let ' s Kick It : How to Stop Wasting the Bottom Third of Your Large - Scale Display. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, (2014).

132. Juckett, G. Cross-cultural medicine. *American Family Physician 72*, 11 (2005), 2267–2274.

133. June, L. For Amusement Only: the life and death of the American arcade. The Verge http://www.theverge.com/2013/1/16/3740422/the-life-and-death-of-the-american-arcade-for-amusement-only, (2013).

134. Kaltenbrunner, M. and Bencina, R. reacTIVision: a computer-vision framework for table-based tangible interaction. *Proc. of the 1st international conference on Tangible and embedded interaction*, ACM (2007), 69–74.

135. Kapur, A., Kapur, A., Virji-Babul, N., Tzanetakis, G., and Driessen, P.F. Gesture-based affective computing on motion capture data. In *Affective Computing and Intelligent Interaction*. Springer, 2005, 1–7.

136. Karam, M. and schrafel, mc. *A taxonomy of gestures in human computer interactions*. University of Southampton, 2005.

137. Karg, M., Samadani, A.-A., Gorbet, R., Kuhnlenz, K., Hoey, J., and Kulic, D. Body movements for affective expression: a survey of automatic recognition and generation. *Affective Computing, IEEE Transactions on 4*, 4 (2013), 341–359.

138. Kerr, Z.Y., Collins, C.L., and Comstock, R.D. Epidemiology of weight training-related injuries presenting to United States emergency departments, 1990 to 2007. *The American Journal of Sports Medicine 38*, 4 (2010), 765–771.

139. Kim, S.-H. and Kaber, D.B. Design and evaluation of dynamic text-editing methods using foot pedals. *International Journal of Industrial Ergonomics 39*, 2 (2009), 358–365.

140. Kipp, M. Anvil-a generic annotation tool for multimodal dialogue. *7th European Conference on Speech Communication and Technology*, (2001).

141. Kirk, D., Rodden, T., and Fraser, D.S. Turn it this way: grounding collaborative action with remote gestures. *Proceedings of the SIGCHI conference on Human factors in computing systems*, ACM (2007), 1039–1048.

142. Kleinsmith, A. and Bianchi-Berthouze, N. Affective Body Expression Perception and Recognition: A Survey. *Affective Computing, IEEE Transactions on PP*, 99 (2012), 1–1.

143. Knight, F., Schwirtz, A., Psomadelis, F., Baber, C., Bristow, W., and Arvanitis, N. The design of the SensVest. *Personal and ubiquitous computing 9*, 1 (2005), 6–19.

144. Kong, K. and Tomizuka, M. A Gait Monitoring System Based on Air Pressure Sensors Embedded in a Shoe. *Mechatronics, IEEE/ASME Transactions on 14*, 3 (2009), 358–370.

145. Koons, D.B., Sparrell, C.J., and Thorisson, K.R. Integrating Simultaneous Input from Speech, Gaze, and Hand Gestures. In M.T. Maybury, ed., *Intelligent Multimedia Interfaces*. American Association for Artificial Intelligence, Menlo Park, CA, USA, 1993, 257–276.

146. Kosunen, I., Jylha, A., Ahmed, I., et al. Comparing eye and gesture pointing to drag items on large screens. *ITS*, ACM (2013), 425–428.

147. Kroemer, K.H.E. Foot operation of controls. *Ergonomics 14*, 3 (1971), 333–361.

148. Krogh, P.G., Ludvigsen, M., and Lykke-olesen, A. Help me pull that cursor - A Collaborative Interactive Floor Enhancing Community Interaction. December (2004), 75–87.

149. Kruyff, P. de, Steentjes, A., and Shahid, S. The Alkaline Arcade: A Child-friendly Fun Machine for Battery Recycling. *Proceedings of the 8th International Conference on Advances in Computer Entertainment Technology*, ACM (2011), 77:1–77:2.

150. Kume, Y., Shirai, A., and Sato, M. Foot Interface : Fantastic Phantom Slipper. *ACM SIGGRAPH 98 Conference abstracts and applications*, ACM (1998), 114—–.

151. Kurillo, G., Bajcsy, R., Nahrsted, K., and Kreylos, O. Immersive 3d environment for remote collaboration and training of physical activities. *Virtual Reality Conference, 2008. VR'08. IEEE*, IEEE (2008), 269–270.

152. Laban, R. von. *Principles of Dance and Movement Notation*. Macdonald & Evans, 1956.

153. Laban, R. von. *Principles of Dance and Movement Notation*. Plays, Incorporated, 1975.

154. LaViola Jr, J.J., Feliz, D.A., Keefe, D.F., and Zeleznik, R.C. Hands-free multi-scale navigation in virtual environments. *Proceedings of the 2001 symposium on Interactive 3D graphics*, ACM (2001), 9–15.

155. Lécuyer, A., Marchal, M., Hamelin, A., et al. Shoes-your-style: changing sound of footsteps to create new walking experiences. *Proceedings of workshop on sound and music computing for human-computer interaction (CHItaly), Alghero, Italy*, (2011), 13–16.

156. Lieberman, H. Your wish is my command: Programming by example. Morgan Kaufmann, 2001.

157. Lippert, L.S. *Clinical kinesiology and anatomy*. FA Davis, 2011.

158. Littlewort, G.C., Bartlett, M.S., and Lee, K. Automatic coding of facial expressions displayed during posed and genuine pain. *Image and Vision Computing 27*, 12 (2009), 1797–1803.

159. Lopes, P.A.S.A., Fernandes, G., and Jorge, J. Trainable DTW-based classifier for recognizing feet-gestures. *16th Portuguese Conference on Pattern Recognition (RecPad 2010)*, (2010).

160. Lovejoy, C.O. Evolution of human walking. *Sci Am 259*, 5 (1988), 118–25.

161. Lövheim, H. A new three-dimensional model for emotions and monoamine neurotransmitters. *Medical hypotheses 78*, 2 (2012), 341–348.

162. Lü, H. and Li, Y. Gesture coder: a tool for programming multi-touch gestures by demonstration. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM (2012), 2875–2884.

163. MacKenzie, C.L. and Iberall, T. *The grasping hand*. Elsevier, 1994.

164. MacKenzie, I.S. Fitts' law as a research and design tool in human-computer interaction. *Human-computer interaction 7*, 1 (1992), 91–139.

165. Mackenzie, I.S. Movement time prediction in human-computer interfaces. *In Readings in Human-Computer Interaction (2nd*, Morgan Kaufmann (1995), 483–493.

166. Mann, S. Smart clothing: The wearable computer and wearcam. *Personal Technologies 1*, 1 (1997), 21–27.

167. Martin, C.C., Burkert, D.C., Choi, K.R., et al. A real-time ergonomic monitoring system using the Microsoft Kinect. *Systems and Information Design Symposium (SIEDS), 2012 IEEE*, IEEE (2012), 50–55.

168. Matthies, D.J.C., Müller, F., Anthes, C., and Kranzlmüller, D. ShoeSoleSense: proof of concept for a wearable foot interface for virtual and real environments. *Proceedings of the 19th ACM Symposium on Virtual Reality Software and Technology - VRST '13*, ACM Press (2013), 93.

169. McElligott, L., Dillon, M., Leydon, K., Richardson, B., Fernström, M., and Paradiso, J.A. ForSe FIElds'-Force Sensors for Interactive Environments. In *UbiComp 2002: Ubiquitous Computing*. Springer, 2002, 168–175.

170. McFarlane, D.C. and Wilder, S.M. Interactive dirt: increasing mobile work performance with a wearable projector-camera system. *Proceedings of the 11th international conference on Ubiquitous computing*, ACM (2009), 205–214.

171. Mehrabian, A. Some referents and measures of nonverbal behavior. *Behavior Research Methods & Instrumentation 1*, 6 (1968), 203–207.

172. Mehrabian, A. Basic dimensions for a general psychological theory: Implications for personality, social, environmental, and developmental studies. Oelgeschlager, Gunn & Hain, Incorporated, 1980.

173. Merlin, B. and Stanislavskij, K. *The complete Stanislavsky toolkit*. Nick Hern Books London, 2007.

174. Michahelles, F. and Schiele, B. Sensing and monitoring professional skiers. *Pervasive Computing, IEEE 4*, 3 (2005), 40–45.

175. Moens, B., Noorden, L. van, and Leman, M. D-Jogger: Syncing music with walking. (2010).

176. Mohamed, F. and Fels, S. LMNKui: Overlaying computer controls on a piano keyboard. *CHI '02 extended abstracts on Human factors in computing systems - CHI '02*, ACM Press (2002), 140.

177. Moller, A., Roalter, L., Diewald, S., et al. Gymskill: A personal trainer for physical exercises. *Pervasive Computing and Communications (PerCom), 2012 IEEE International Conference on*, IEEE (2012), 213–220.

178. Morley Jr, R.E., Richter, E.J., Klaesner, J.W., Maluf, K.S., Mueller, M.J., and Morley, R.E. In-shoe multisensory data acquisition system. *Biomedical Engineering, IEEE Transactions on 48*, 7 (2001), 815–820.

179. Morris, D. *Peoplewatching*. Random House, 2002.

180. Mueller, F., Agamanolis, S., and Picard, R. Exertion interfaces: sports over a distance for social bonding and fun. *Proceedings of the SIGCHI conference on Human factors in computing systems*, ACM (2003), 561–568.

181. Mullane, S., Chakravorti, N., Conway, P., and West, A. Design and implementation of a user-centric swimming performance monitoring tool. *Proceedings of the Institution of Mechanical Engineers, Part P: Journal of Sports Engineering and Technology 225*, (2011), 213–229.

182. Natapov, D., Castellucci, S.J., and MacKenzie, I.S. ISO 9241-9 evaluation of video game controllers. *Proceedings of Graphics Interface 2009*, Canadian Information Processing Society (2009), 223–230.

183. Navarro, J. and Karlins, M. *What every body is saying*. HarperCollins, 2009.

184. Nguyen, Q. and Kipp, M. Annotation of human gesture using 3d skeleton controls. *Proc. of the Seventh International Conference on Language Resources and Evaluation, LREC*, Citeseer (2010).

185. Nilsson, N.C., Serafin, S., Laursen, M.H., Pedersen, K.S., Sikstrom, E., and Nordahl, R. Tapping-in-place: Increasing the naturalness of immersive walking-in-place locomotion through novel gestural input. *3D User Interfaces (3DUI), 2013 IEEE Symposium on*, IEEE (2013), 31–38.

186. Norman, D.A. and Nielsen, J. Gestural interfaces: a step backward in usability. *interactions 17*, 5 (2010), 46–49.

187. Nundy, S., Lotto, B., Coppola, D., Shimpi, A., and Purves, D. Why are angles misperceived? *Proceedings of the National Academy of Sciences 97*, 10 (2000), 5592–5597.

188. O'Donovan, G., Blazevich, A.J., Boreham, C., et al. The ABC of Physical Activity for Health: a consensus statement from the British Association of Sport and Exercise Sciences. *Journal of sports sciences 28*, 6 (2010), 573–591.

189. Ohtaki, Y., Susumago, M., Suzuki, A., Sagawa, K., Nagatomi, R., and Inooka, H. Automatic classification of ambulatory movements and evaluation of energy consumptions utilizing accelerometers and a barometer. *Microsystem technologies 11*, 8-10 (2005), 1034–1040.

190. Oliver, N. and Flores-Mangas, F. MPTrain: a mobile, music and physiology-based personal trainer. *Proceedings of the 8th conference on Human-computer interaction with mobile devices and services*, ACM (2006), 21–28.

191. Orecchini, G., Yang, L., Tentzeris, M., and Roselli, L. Smart Shoe": An autonomous inkjet-printed RFID system scavenging walking energy. *Antennas and Propagation (APSURSI), 2011 IEEE International Symposium on*, IEEE (2011), 1417–1420.

192. Orr, R.J. and Abowd, G.D. The smart floor: a mechanism for natural user identification and tracking. *CHI'00 extended abstracts on Human factors in computing systems*, ACM (2000), 275–276.

193. Paelke, V., Reimann, C., and Stichling, D. Foot-based mobile interaction with games. *Proceedings of the 2004 ACM SIGCHI International Conference on Advances in computer entertainment technology*, ACM (2004), 321–324.

194. Paillard, J. Le corps situé et le corps identifié. *Rev. Méd. Suisse Romande 100*, 129.141 (1980).

195. Pakkanen, T. and Raisamo, R. Appropriateness of foot interaction for non-accurate spatial tasks. *CHI'04 extended abstracts on Human factors in computing systems*, ACM (2004), 1123–1126.

196. Papetti, S., Civolani, M., and Fontana, F. Rhythm ' n ' Shoes : a wearable foot tapping interface with audio-tactile feedback. *Proceedings of international conference of new interfaces for musical expression*, (2011), 473–476.

197. Paradiso, J., Abler, C., Hsiao, K., and Reynolds, M. The magic carpet: physical sensing for immersive environments. *CHI'97 Extended Abstracts on Human Factors in Computing Systems*, ACM (1997), 277–278.

198. Paradiso, J.A., Hsiao, K., Benbasat, A.Y., and Teegarden, Z. Design and implementation of expressive footwear. *IBM Systems Journal 39*, 3.4 (2000), 511–529.

199. Paradiso, J.A., Hsiao, K.-Y., and Hu, E. Interactive music for instrumented dancing shoes. *Proc. of the 1999 International Computer Music Conference*, Citeseer (1999), 453–456.

200. Paradiso, J.A., Hu, E., and Hsiao, K. Instrumented footwear for interactive dance. *Proc. of the XII Colloquium on Musical Informatics*, (1998), 24–26.

201. Paradiso, J.A., Hu, E., and Hsiao, K.Y. The CyberShoe: a wireless multisensor interface for a dancer's feet. *Proceedings of International Dance and Technology 99*, (1999), 57–60.

202. Paradiso, J.A. and Hu, E. Expressive footwear for computer-augmented dance performance. *Wearable Computers, 1997. Digest of Papers., First International Symposium on*, IEEE (1997), 165–166.

203. Paradiso, J.A., Morris, S.J., Benbasat, A.Y., and Asmussen, E. Interactive therapy with instrumented footwear. *CHI'04 Extended Abstracts on Human Factors in Computing Systems*, ACM (2004), 1341–1343.

204. Parrott, W.G. Emotions in social psychology: Essential readings. Psychology Press, 2001.

205. Pasch, M., Bianchi-Berthouze, N., Dijk, B. van, and Nijholt, A. Movement-based sports video games: Investigating motivation and gaming experience. *Entertainment Computing 1*, 2 (2009), 49–61.

206. Pearson, G. and Weiser, M. Of moles and men: the design of foot controls for workstations. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM (1986), 333–339.

207. Pearson, G. and Weiser, M. Exploratory evaluation of a planar foot-operated cursor-positioning device. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM (1988), 13–18.

208. Pearson, G. and Weiser, M. Exploratory Evaluations of two Versions of a Foot-Operated Cursor-Positioning Device in a Target-Selection Task. *ACM SIGCHI Bulletin 19*, 3 (1988), 70–75.

209. Picard, R.W. *Affective computing*. MIT press, 2000.

210. Plutchik, R. *Emotion: A psychoevolutionary synthesis*. Harpercollins College Division, 1980.

211. Pouke, M., Karhu, A., Hickey, S., and Arhippainen, L. Gaze tracking and non-touch gesture based interaction method for mobile 3D virtual spaces. *OzCHI*, ACM (2012), 505–512.

212. Poupyrev, I., Billinghurst, M., Weghorst, S., and Ichikawa, T. The Go-go Interaction Technique: Non-linear Mapping for Direct Manipulation in VR. *Proceedings of the 9th Annual ACM Symposium on User Interface Software and Technology*, ACM (1996), 79–80.

213. Powers, S. Build Your Own Arcade Game Player and Relive the '80s! *Linux J. 2007*, 160 (2007), 1–.

214. Ptaszynski, M., Maciejewski, J., Dybala, P., Rzepka, R., and Araki, K. Cao: A fully automatic emoticon analysis system based on theory of kinesics. *Affective Computing, IEEE Transactions on 1*, 1 (2010), 46–59.

215. Quek, F., Ehrich, R., and Lockhart, T. As go the feet...: on the estimation of attentional focus from stance. *Proceedings of the 10th international conference on Multimodal interfaces*, ACM (2008), 97–104.

216. Quek, F., McNeill, D., Bryll, R., et al. Multimodal human discourse: gesture and speech. *ACM Transactions on Computer-Human Interaction (TOCHI) 9*, 3 (2002), 171–193.

217. Raffle, H. Topobo: programming by example to create complex behaviors. *Proceedings of the 9th International Conference of the Learning Sciences-Volume 2*, International Society of the Learning Sciences (2010), 126–127.

218. Ramadoss, B. and Rajkumar, K. Semi-automated annotation and retrieval of dance media objects. *Cybernetics and Systems: An International Journal 38*, 4 (2007), 349–379.

219. Rehman, S. ur, Khan, A., Li, H., and Physics, A. Interactive Feet for Mobile Immersive Interaction. *MobileHCI 2012 : Mobile Vision (MobiVis) – Vision-based Applications and HCI*, MOBIVIS (2012).

220. Reiss, A., Hendeby, G., Bleser, G., and Stricker, D. Activity recognition using biomechanical model based pose estimation. In *Smart Sensing and Context*. Springer, 2010, 42–55.

221. Richardson, B., Leydon, K., Fernstrom, M., and Paradiso, J.A. Z-Tiles: building blocks for modular, pressure-sensing floorspaces. *CHI'04 extended abstracts on Human factors in computing systems*, ACM (2004), 1529–1532.

222. Richter, S., Holz, C., and Baudisch, P. Bootstrapper: recognizing tabletop users by their shoes. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM (2012), 1249–1252.

223. Roaas, A. and Andersson, G.B. Normal range of motion of the hip, knee and ankle joints in male subjects, 30-40 years of age. *Acta Orthopaedica 53*, 2 (1982), 205–208.

224. Rosenblum, S.P. Pedaling the piano: A brief survey from the eighteenth century to the present. *Performance Practice Review 6*, 2 (1993), 8.

225. Rovers, A.F. and Van Essen, H.A. FootIO - design and evaluation of a device to enable foot interaction over a computer network. *Eurohaptics Conference, 2005 and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems, 2005. World Haptics 2005. First Joint*, (2005), 521–522.

226. Rovers, A.F. and Van Essen, H.A. Guidelines for haptic interpersonal communication applications: an exploration of foot interaction styles. *Virtual Reality 9*, 2-3 (2006), 177–191.

227. Russell, J.A. A circumplex model of affect. *Journal of personality and social psychology 39*, 6 (1980), 1161.

228. Saffer, D. Designing gestural interfaces: Touchscreens and interactive devices. O'Reilly Media, Inc., 2008.

229. Samadani, A.-A., Burton, S., Gorbet, R., and Kulic, D. Laban effort and shape analysis of affective hand and arm movements. *Affective Computing and Intelligent Interaction (ACII), 2013 Humaine Association Conference on*, IEEE (2013), 343–348.

230. Sangsuriyachot, N., Mi, H., and Sugimoto, M. Novel interaction techniques by combining hand and foot gestures on tabletop environments. *Proc. ITS '11*, ACM (2011), 268–269.

231. Scherer, K.R. Psychological models of emotion. *The neuropsychology of emotion 137*, 3 (2000), 137–162.

232. Schmidt, A. Implicit human computer interaction through context. *Personal technologies 4*, 2-3 (2000), 191–199.

233. Schmidt, D., Ramakers, R., Pedersen, E., et al. Kickables: Tangibles for Feet. *CHI'14*, (2014).

234. Schmidt, T., Duncan, S., Ehmer, O., et al. An exchange format for multimodal annotations. In *Multimodal corpora*. Springer, 2009, 207–221.

235. Schöning, J., Daiber, F., Krüger, A., and Rohs, M. Using hands and feet to navigate and manipulate spatial data. *CHI '09 Extended Abstracts on Human Factors in Computing Systems*, ACM (2009), 4663–4668.

236. Scott, J., Dearman, D., Yatani, K., and Truong, K.N. Sensing foot gestures from the pocket. *Proceedings of the 23nd annual ACM symposium on User interface software and technology*, ACM (2010), 199–208.

237. Scott, J., Dearman, D., Yatani, K., and Truong, K.N. Sensing foot gestures from the pocket. *Proceedings of the 23nd annual ACM symposium on User interface software and technology*, ACM (2010), 199–208.

238. Sellen, A.J., Kurtenbach, G.P., and Buxton, W.A. The prevention of mode errors through sensory feedback. *Human-Computer Interaction 7*, 2 (1992), 141–164.

239. Shaukat, S., Yousaf, M.H., and Habib, H.A. Real-time feet movement detection and tracking for controlling a Toy car. *Advanced Computer Theory and Engineering (ICACTE), 2010 3rd International Conference on*, IEEE (2010), V6—–5.

240. Sibert, L.E. and Jacob, R.J. Evaluation of eye gaze interaction. *Proceedings of the SIGCHI conference on Human factors in computing systems*, ACM (2000), 281–288.

241. Simeone, A. and Gellersen, H. Comparing Direct and Indirect Touch in a Stereoscopic Interaction Task. *3D User Interfaces (3DUI), 2015 IEEE Symposium on*, (2015).

242. Simeone, A., Velloso, E., Alexander, J., and Gellersen, H. Feet Movement in Desktop 3D Interaction. *Proceedings of the 2014 IEEE Symposium on 3D User Interfaces*, IEEE (2014).

243. Simonen, R.L., Videman, T., Battié, M.C., and Gibbons, L.E. Comparison of foot and hand reaction times among men: A methodologic study using simple and multiple-choice repeated measurements. *Perceptual and Motor Skills*, (1995).

244. Skoglund, A., Iliev, B., and Palm, R. Programming-by-Demonstration of reaching motions—A next-state-planner approach. *Robotics and Autonomous Systems 58*, 5 (2010), 607–621.

245. Slater, M., Usoh, M., and Steed, A. Taking steps: the influence of a walking technique on presence in virtual reality. *ACM Transactions on Computer-Human Interaction (TOCHI) 2*, 3 (1995), 201–219.

246. Sørensen, R.K. Project MAME - Build your own MAME cabinet. 2008.

247. Soukoreff, R.W. and MacKenzie, I.S. Towards a standard for pointing device evaluation, perspectives on 27 years of Fitts' law research in HCI. *International journal of human-computer studies 61*, 6 (2004), 751–789.

248. Spina, G., Huang, G., Vaes, A., Spruit, M., and Amft, O. COPDTrainer: a smartphone-based motion rehabilitation training system with real-time acoustic feedback. *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing*, ACM (2013), 597–606.

249. Springer, J. and Siebes, C. Position controlled input device for handicapped: Experimental studies with a footmouse. *International Journal of Industrial Ergonomics 17*, 2 (1996), 135–152.

250. Standard, E. ISO 9000: 2005. Quality management system-Fundamentals and vocabulary, ISO 1, (2005), 1.

251. Stellmach, S. and Dachselt, R. Investigating gaze-supported multimodal pan and zoom. *Proceedings of the Symposium on Eye Tracking Research and Applications*, ACM (2012), 357–360.

252. Stellmach, S. and Dachselt, R. Look & touch: gaze-supported target acquisition. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM (2012), 2981–2990.

253. Stellmach, S. and Dachselt, R. Still looking: Investigating seamless gaze-supported selection, positioning, and manipulation of distant targets. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM (2013), 285–294.

254. Stellmach, S., Stober, S., Nürnberger, A., and Dachselt, R. Designing gaze-supported multimodal interactions for the exploration of large image collections. *Proceedings of the 1st Conference on Novel Gaze-Controlled Applications*, ACM (2011), 1.

255. Stellmach, S. Gaze-supported Multimodal Interaction. 2013. http://www.dr.hut-verlag.de/978-3-8439-1235-8.html.

256. Stienstra, J., Overbeeke, K., and Wensveen, S. Embodying complexity through movement sonification: case study on empowering the speed-skater. *Proceedings of the 9th ACM SIGCHI Italian Chapter International Conference on Computer-Human Interaction: Facing Complexity*, ACM (2011), 39–44.

257. Strohrmann, C., Harms, H., Tröster, G., Hensler, S., and Müller, R. Out of the lab and into the woods: kinematic analysis in running using wearable sensors. *Proceedings of the 13th international conference on Ubiquitous computing*, ACM (2011), 119–122.

258. Sundstedt, V. Gazing at games: using eye tracking to control virtual characters. *ACM SIGGRAPH 2010 Courses*, ACM (2010), 5.

259. Sykes, J. and Brown, S. Affective gaming: measuring emotion through the gamepad. *CHI'03 extended abstracts on Human factors in computing systems*, ACM (2003), 732–733.

260. Tang, R., Alizadeh, H., Tang, A., Bateman, S., and Jorge, J.A.P. Physio@Home: Design Explorations to Support Movement Guidance. *CHI '14 Extended Abstracts on Human Factors in Computing Systems*, ACM (2014), 1651–1656.

261. Tanriverdi, V. and Jacob, R.J. Interacting with eye movements in virtual environments. *Proceedings of the SIGCHI conference on Human factors in computing systems*, ACM (2000), 265–272.

262. Tao, Y., Lam, T.L., Qian, H., and Xu, Y. A real-time intelligent shoe-keyboard for computer input. *2012 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, IEEE (2012), 1488–1493.

263. Tapia, E.M., Intille, S.S., and Larson, K. Activity recognition in the home using simple and ubiquitous sensors. Springer, 2004.

264. Templeman, J.N., Denbrook, P.S., and Sibert, L.E. Virtual locomotion: Walking in place through virtual environments. *Presence: Teleoperators and Virtual Environments 8*, 6 (1999), 598–617.

265. Thorp, E.O. The invention of the first wearable computer. *Second International Symposium on Wearable Computers*, (1998), 4—–8.

266. Tian, Y.I., Kanade, T., and Cohn, J.F. Recognizing action units for facial expression analysis. *Pattern Analysis and Machine Intelligence, IEEE Transactions on 23*, 2 (2001), 97–115.

267. Tomkins, S. Affect/imagery/consciousness. Vol. 2: The negative affects. (1963).

268. Tomkins, S.S. Affect, imagery, consciousness: Vol. I. The positive affects. (1962).

269. Tran, C., Doshi, A., and Trivedi, M.M. Modeling and prediction of driver behavior by foot gesture analysis. *Computer Vision and Image Understanding 116*, 3 (2012), 435–445.

270. Trombley, D.J. *Experimental Determination of an Optimal Foot Pedal Design*. Central Library Texas, Technological College, Lubbock (TX), 1966.

271. Turner, J., Alexander, J., Bulling, A., Schmidt, D., and Gellersen, H. Eye pull, eye push: Moving objects between large screens and personal devices with gaze and touch. In *Human-Computer Interaction–INTERACT 2013*. Springer, 2013, 170–186.

272. Turner, J., Velloso, E., Gellersen, H., and Sundstedt, V. EyePlay: applications for gaze in games. *Proceedings of the first ACM SIGCHI annual symposium on Computer-human interaction in play*, ACM (2014), 465–468.

273. Velloso, E., Alexander, J., Bulling, A., and Gellersen, H. Interactions Under the Desk: A Characterisation of Foot Movements for Input in a Seated Position. *Proc. of the 15th IFIP TC13 Conference on Human-Computer Interaction (INTERACT 2015)*, (2015).

274. Velloso, E., Bulling, A., Gellersen, H., Ugulino, W., and Fuks, H. Qualitative activity recognition of weight lifting exercises. *Proceedings of the 4th Augmented Human International Conference*, ACM (2013), 116–123.

275. Velloso, E., Bulling, A., and Gellersen, H. Towards qualitative assessment of weight lifting exercises using body-worn sensors. *Proceedings of the 13th international conference on Ubiquitous computing*, ACM (2011), 587–588.

276. Velloso, E., Bulling, A., and Gellersen, H. AutoBAP: Automatic coding of body action and posture units from wearable sensors. *Affective Computing and Intelligent Interaction (ACII), 2013 Humaine Association Conference on*, IEEE (2013), 135–140.

277. Velloso, E., Bulling, A., and Gellersen, H. MotionMA: Motion Modelling and Analysis by Demonstration. *Proc. of the 31st SIGCHI International Conference on Human Factors in Computing Systems*, (2013).

278. Velloso, E., Cardador, D., Vega, K., et al. The Web of Things as an Infrastructure for Improving Users' Health and Wellbeing. *II Workshop of the Brazilian Institute for Web Science Research, Rio de Janeiro, Brazil*, (2011).

279. Velloso, E., Turner, J., Alexander, J., Bulling, A., and Gellersen, H. An Empirical Investigation of Gaze Selection in Mid-Air Gestural 3D Manipulation. *Proc. of the 15th IFIP TC13 Conference on Human-Computer Interaction (INTERACT 2015)*, (2015).

280. Wagner, J., Nancel, M., Gustafson, S.G., Huot, S., and Mackay, W.E. Body-centric design space for multi-surface interaction. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '13*, ACM Press (2013), 1299.

281. Watanabe, J., Ando, H., and Maeda, T. Shoe-shaped interface for inducing a walking cycle. *Proceedings of the 2005 international conference on Augmented tele-existence*, ACM (2005), 30–34.

282. Weaver, T.D. and Klein, R. The evolution of human walking. *Human Walking*, (2006), 23–32.

283. Wendt, J.D., Whitton, M.C., and Brooks Jr, F.P. Gud wip: Gait-understanding-driven walking-in-place. *Virtual Reality Conference (VR), 2010 IEEE*, IEEE (2010), 51–58.

284. Williams, B., Bailey, S., Narasimham, G., Li, M., and Bodenheimer, B. Evaluation of walking in place on a wii balance board to explore a virtual environment. *ACM Transactions on Applied Perception (TAP) 8*, 3 (2011), 19.

285. Williams, L., Groves, D., and Thurgood, G. *Strength Training: The Complete step-by-step guide to sculpting a stronger body*. Dorling Kinderley Limited,, 2011.

286. Wobbrock, J.O., Shinohara, K., and Jansen, A. The effects of task dimensionality, endpoint deviation, throughput calculation, and experiment design on pointing measures and models. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM (2011), 1639–1648.

287. Xavier, E., Filho, D.L., Nunes, M.B., et al. Why Not with the Foot? *2011 Brazilian Symposium on Games and Digital Entertainment*, IEEE (2011), 270–281.

288. Yamamoto, T., Tsukamoto, M., and Yoshihisa, T. Foot-Step Input Method for Operating Information Devices While Jogging. *Applications and the Internet, 2008. SAINT 2008. International Symposium on*, (2008), 173–176.

289. Yamauchi, T. Mouse Trajectories and State Anxiety: Feature Selection with Random Forest. *Affective Computing and Intelligent Interaction (ACII), 2013 Humaine Association Conference on*, IEEE (2013), 399–404.

290. Ye, W., Xu, Y., and Lee, K.K. Shoe-Mouse: An integrated intelligent shoe. *Intelligent Robots and Systems, 2005.(IROS 2005). 2005 IEEE/RSJ International Conference on*, IEEE (2005), 1163–1167.

291. Yoo, B., Han, J.-J., Choi, C., et al. 3D user interface combining gaze and hand gestures for large-scale display. *CHI'10 Extended Abstracts on Human Factors in Computing Systems*, ACM (2010), 3709–3714.

292. Yousaf, M.H. and Habib, H.A. Adjustable Pedals by Gesture Recognition: A Novel Approach for User Interface in Automotive. *Middle-East Journal of Scientific Research 11*, 11 (2012), 1575–1583.

293. Zeng, Z., Pantic, M., Roisman, G.I., and Huang, T.S. A survey of affect recognition methods: Audio, visual, and spontaneous expressions. *Pattern Analysis and Machine Intelligence, IEEE Transactions on 31*, 1 (2009), 39–58.

294. Zhai, S., Morimoto, C., and Ihde, S. Manual and gaze input cascaded (MAGIC) pointing. *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, ACM (1999), 246–253.

295. Zheng, Y., McCaleb, M., Strachan, C., and Williams, B. Exploring a virtual environment by walking in place using the Microsoft Kinect. *Proceedings of the ACM symposium on applied perception*, ACM (2012), 131–131.

296. Zhong, K., Tian, F., and Wang, H. Foot Menu: using heel rotation information for menu selection. *Wearable Computers (ISWC), 2011 15th Annual International Symposium on*, IEEE (2011), 115–116.

297. Zinnen, A., Blanke, U., and Schiele, B. An analysis of sensor-oriented vs. model-based activity recognition. *Wearable Computers, 2009. ISWC'09. International Symposium on*, IEEE (2009), 93–100.

298. Zinnen, A., Wojek, C., and Schiele, B. Multi activity recognition based on bodymodel-derived primitives. In *Location and Context Awareness*. Springer, 2009, 1–18.

9 Supplementary Material

Appendix 1 - Pre Questionnaire for the study in Section 2.4.6

QUESTIONNAIRE

All the information provided by you is strictly confidential.

Name: _____

Date of birth: _____/_____/_____

Today's date: _____/_____/_____

Current time: _____:_____

Height: _____ **(meters)**

Weight: _____ **(kg)**

Forearm length: _____ **(cm)**

Arm length: _____ **(cm)**

Waistline: _____ **(cm)**

What kinds of physical activities do you currently do?

_____

_____

_____

Appendix 2 - Post-study questionnaire for the study in Section 2.4.6

POST EXPERIMENT QUESTIONNAIRE

Did you feel any difference in performing the exercise with and without the feedback system?

How would you qualify the visual feedback provided by the system?

How helpful do you think a system such as this is in a gym environment?

Not helpful at all 1 2 3 4 5 Very helpful

How clear was the presentation of information?

Not clear at all 1 2 3 4 5 Very clear

How much do you believe the feedback influenced your performance?

Not at all 1 2 3 4 5 A lot

How much you agree that you tried to change your movements according to the feedback?

Strongly disagree 1 2 3 4 5 Strongly agree

How much do you agree that the feedback improved your performance?

Strongly disagree 1 2 3 4 5 Strongly agree

Any other comments?

Appendix 3 - Pre-Study Questionnaire for the study in Section 2.4.7

Questionnaire

Name:

Date of birth:

Gender:

How would you rate your experience with weight lifting?

- **Expert**
- **Very Experienced**
- **Experienced**
- **Little Experience**
- **No experience**

Approximately how many years of experience have you got with weight lifting?

Have you got formal training in weight lifting/fitness instruction/physical education? If yes, please specify.

How familiar are you with the following exercises?

Unilateral Dumbbell Biceps Curl

- **I can perform it and instruct other on how to perform it.**
- **I can perform it, but wouldn't feel comfortable instructing others on how to perform it.**
- **I can't perform it, but I can instruct others on how to perform it.**
- **I can't perform it and wouldn't feel comfortable instructing others on how to perform it.**

Unilateral Dumbbell I Lateral Raise

- **I can perform it and instruct other on how to perform it.**
- **I can perform it, but wouldn't feel comfortable instructing others on how to perform it.**
- **I can't perform it, but I can instruct others on how to perform it.**
- **I can't perform it and wouldn't feel comfortable instructing others on how to perform it.**

Unilateral Dumbbell Triceps Extension

- **I can perform it and instruct other on how to perform it.**
- **I can perform it, but wouldn't feel comfortable instructing others on how to perform it.**
- **I can't perform it, but I can instruct others on how to perform it.**
- **I can't perform it and wouldn't feel comfortable instructing others on how to perform it.**

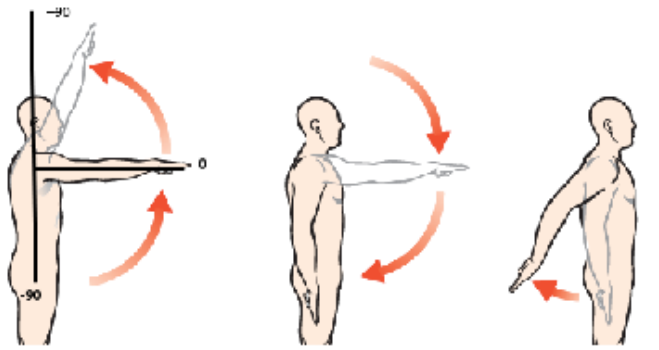
Appendix 4 – Parameter estimation questionnaire for the study in Section 2.5.7

Exercise Modelling Questionnaire

In this questionnaire, you will be asked to fill in an estimate of the acceptable ranges of motion for certain joint movements for three separate exercises.

Name:

Unilateral Biceps Curl



Flexion Extension Hyperextension

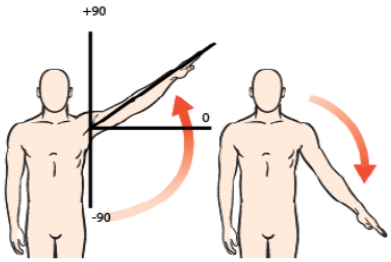
What is the initial angle?

Is there shoulder flexion/extension/hyperextension in this exercise?

If no, what is the acceptable range?

If yes, what is the angle in the end of the motion?

If yes, how long should this movement take?



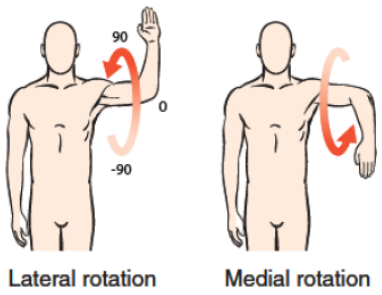
What is the initial angle?

Is there shoulder abduction/adduction in this exercise?

If no, what is the acceptable range?

If yes, what is the angle in the end of the motion?

If yes, how long should this movement take?



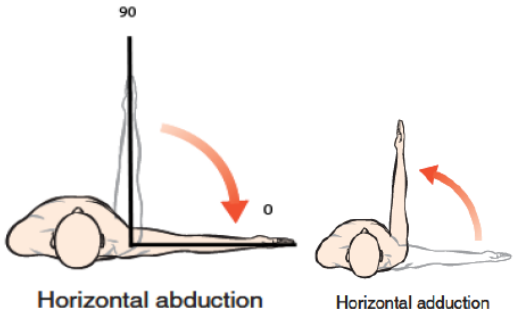
What is the initial angle?

Is there shoulder lateral/medial rotation in this exercise?

If no, what is the acceptable range?

If yes, what is the angle in the end of the motion?

If yes, how long should this movement take?



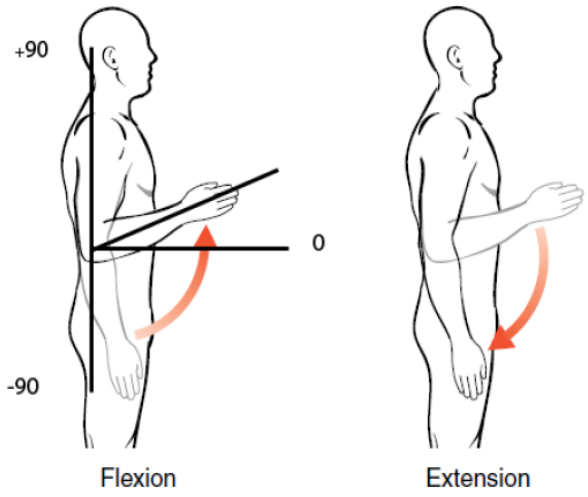
What is the initial angle?

Is there shoulder horizontal abduction/adduction in this exercise?

If no, what is the acceptable range?

If yes, what is the angle in the end of the motion?

If yes, how long should this movement take?



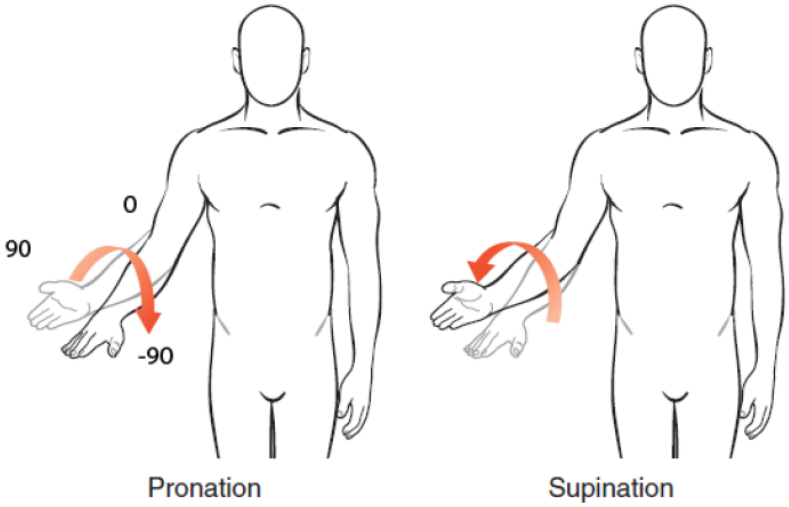
What is the initial angle?

Is there elbow flexion/extension in this exercise?

If no, what is the acceptable range?

If yes, what is the angle in the end of the motion?

If yes, how long should this movement take?



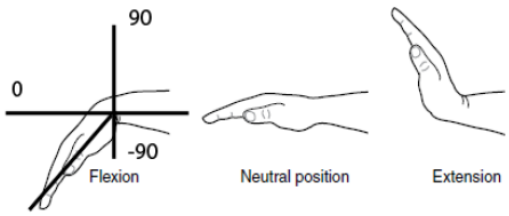
What is the initial angle?

Is there forearm pronation/supination in this exercise?

If no, what is the acceptable range?

If yes, what is the angle in the end of the motion?

If yes, how long should this movement take?



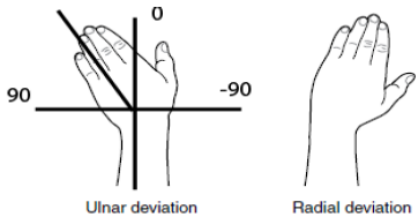
What is the initial angle?

Is there wrist flexion/extension in this exercise?

If no, what is the acceptable range?

If yes, what is the angle in the end of the motion?

If yes, how long should this movement take?



What is the initial angle?

Is there ulnar/radial deviation in this exercise?

If no, what is the acceptable range?

If yes, what is the angle in the end of the motion?

If yes, how long should this movement take?

Appendix 5 - Pre-study questionnaire for the study in Section 2.5.4

Personal Questionnaire

Name:

Age:

Gender: M F

Height:

Weight:

How experienced are you with weight lifting:

No experience 1 2 3 4 5 **Expert**

Are you right- or left-handed?

Please check which of the following video games have you ever played:

- Wii Fit
- Nike+ Kinect Training
- Your Shape
- EA Sports Active
- Other exercise-related video game:_____

Appendix 6 - Questionnaire in which participants described movements and evaluated the model extracted for it in Section 2.5.4

Demonstration Evaluation

Name: _____

Movement Definition

Think of a **body movement** that you would wish to teach someone else. This movement should be repeatable and controlled. Please describe **every step** in its execution in as much detail as possible.

_____

_____

_____

_____

_____

_____

What would you **call** this movement?

_____

Describe **five mistakes** that someone else trying to repeat this exact movement would most likely make?

1_____

2_____

3_____

4_____

5_____

How long should each repetition take?_____

Movement Model

For each of the bones, please indicate how **accurate** the system was at modelling its movement and whether it marked it correctly as **dynamic or static**:

Right Arm							
	Polar	1	2	3	4	5	D/S Right	D/S Wrong
	Azimuth	1	2	3	4	5	D/S Right	D/S Wrong
Left Arm							
	Polar	1	2	3	4	5	D/S Right	D/S Wrong
	Azimuth	1	2	3	4	5	D/S Right	D/S Wrong
Right Forearm							
	Polar	1	2	3	4	5	D/S Right	D/S Wrong
	Azimuth	1	2	3	4	5	D/S Right	D/S Wrong
Left Forearm							
	Polar	1	2	3	4	5	D/S Right	D/S Wrong
	Azimuth	1	2	3	4	5	D/S Right	D/S Wrong
Right Hand							
	Polar	1	2	3	4	5	D/S Right	D/S Wrong
	Azimuth	1	2	3	4	5	D/S Right	D/S Wrong
Left Hand							
	Polar	1	2	3	4	5	D/S Right	D/S Wrong
	Azimuth	1	2	3	4	5	D/S Right	D/S Wrong
Right Shoulder							
	Polar	1	2	3	4	5	D/S Right	D/S Wrong
	Azimuth	1	2	3	4	5	D/S Right	D/S Wrong
Left Shoulder							
	Polar	1	2	3	4	5	D/S Right	D/S Wrong
	Azimuth	1	2	3	4	5	D/S Right	D/S Wrong

Column							
	Polar	1	2	3	4	5	D/S Right	D/S Wrong
	Azimuth	1	2	3	4	5	D/S Right	D/S Wrong
Right Leg							
	Polar	1	2	3	4	5	D/S Right	D/S Wrong
	Azimuth	1	2	3	4	5	D/S Right	D/S Wrong
Left Leg							
	Polar	1	2	3	4	5	D/S Right	D/S Wrong
	Azimuth	1	2	3	4	5	D/S Right	D/S Wrong
Right Shin							
	Polar	1	2	3	4	5	D/S Right	D/S Wrong
	Azimuth	1	2	3	4	5	D/S Right	D/S Wrong
Left Shin							
	Polar	1	2	3	4	5	D/S Right	D/S Wrong
	Azimuth	1	2	3	4	5	D/S Right	D/S Wrong
Neck							
	Polar	1	2	3	4	5	D/S Right	D/S Wrong
	Azimuth	1	2	3	4	5	D/S Right	D/S Wrong
Right Foot							
	Polar	1	2	3	4	5	D/S Right	D/S Wrong
	Azimuth	1	2	3	4	5	D/S Right	D/S Wrong
Left Foot							
	Polar	1	2	3	4	5	D/S Right	D/S Wrong
	Azimuth	1	2	3	4	5	D/S Right	D/S Wrong
Left Hip							
	Polar	1	2	3	4	5	D/S Right	D/S Wrong
	Azimuth	1	2	3	4	5	D/S Right	D/S Wrong

Right Hip

Polar	1	2	3	4	5	D/S Right	D/S Wrong
Azimuth	1	2	3	4	5	D/S Right	D/S Wrong

Lower Back

Polar	1	2	3	4	5	D/S Right	D/S Wrong
Azimuth	1	2	3	4	5	D/S Right	D/S Wrong

Which bones' movements could be used to count repetitions?

_____
_____
_____
_____
_____

How good was the choice of the bone to count repetitions?

Very good 1 2 3 4 5 Very bad

Feedback Interface

While performing the **correct** movement, how accurate was the system at:

Counting repetitions?

Very accurate 1 2 3 4 5 Very inaccurate

Displaying the correct **range of motion** of dynamic bones in the dials?

Very accurate 1 2 3 4 5 Very inaccurate

Showing a **green light** for static bones in the traffic lights?

Very accurate 1 2 3 4 5 Very inaccurate

Indicating the correct **speed**?

Very accurate 1 2 3 4 5 Very inaccurate

While performing the **variations** of the movement, how accurately did the system spot each one of them?

Variation 1

Very accurately 1 2 3 4 5 Very inaccurately

Variation 2

Very accurately 1 2 3 4 5 Very inaccurately

Variation 3

Very accurately 1 2 3 4 5 Very inaccurately

Variation 4

Very accurately 1 2 3 4 5 Very inaccurately

Variation 5

Very accurately 1 2 3 4 5 Very inaccurately

Entire System

How much would you agree with the following statements?

The system was able to extract an accurate model of the movement I demonstrated.

Strongly agree 1 2 3 4 5 strongly disagree

The system was able to detect a correct performance of the movement.

Strongly agree 1 2 3 4 5 strongly disagree

The system was able to count repetitions of the movement accurately.

Strongly agree 1 2 3 4 5 strongly disagree

The system was able to detect variations of the movement.

Strongly agree 1 2 3 4 5 strongly disagree

Please give some comments about the system in general:

_____
_____
_____
_____
_____

Appendix 7 - Pre-Study Questionnaire for the study in Section 5.1

Movement Times of Foot Interaction

*Required

1. **Participant ID** *
   The researcher will assign you a participant ID

   _____

2. **Age** *

   _____

3. **Foot size in cm** *
   The researcher will provide you with a tape measure.

   _____

4. **Are you a regular driver?** *
   Mark only one oval.

   ◯ Yes
   ◯ No

5. **What is your dominant hand?** *
   i.e. the hand you use to write
   Mark only one oval.

   ◯ Right
   ◯ Left
   ◯ Both

6. **What is your dominant foot?** *
   i.e. when you climb a step, which foot goes first? when you are pushed from behind, which foot to you put forward to regain balance?
   Mark only one oval.

   ◯ Right
   ◯ Left
   ◯ Both

7. **How experienced are you with foot-operated computer interfaces?**
e.g. foot mouse, foot trackball, pedals
Mark only one oval.

	1	2	3	4	5	
No experience	◯	◯	◯	◯	◯	Expert

Appendix 8 - Post-Study Questionnaire for the study in Section 5.1

Device Acessment Questionnaire

*Required

1. **Participant ID** *

2. **The smoothness during the operation was:** *
Mark only one oval.

	1	2	3	4	5	
Very rough	◯	◯	◯	◯	◯	Very smooth

3. **The mental effort required for the operation was:** *
Mark only one oval.

	1	2	3	4	5	
Very Low	◯	◯	◯	◯	◯	Very High

4. **The physical effort required for operation was** *
Mark only one oval.

	1	2	3	4	5	
Very Low	◯	◯	◯	◯	◯	Very High

5. **Foot fatigue** *
Mark only one oval.

	1	2	3	4	5	
None	◯	◯	◯	◯	◯	Very High

6. **Leg fatigue** *
Mark only one oval.

	1	2	3	4	5	
None	◯	◯	◯	◯	◯	Very High

7. **Thigh fatigue** *
Mark only one oval.

	1	2	3	4	5	
None	◯	◯	◯	◯	◯	Very High

8. **General comfort** *
Mark only one oval.

	1	2	3	4	5	
Very uncomfortable	◯	◯	◯	◯	◯	Very comfortable

9. **Overall the foot tracker was** *
Mark only one oval.

	1	2	3	4	5	
Very difficult to use	◯	◯	◯	◯	◯	Very easy to use

Appendix 9 - Pre-Study Questionnaire for the study in Section 5.2

Edit this

Feet Movement Times

*Required

Name *

Age

Foot size

Do you drive? *
- Yes
- No

What is your dominant hand? *
e.g. the hand you use to write
- Right
- Left
- Both

What is your dominant foot? *
e.g. when you climb a step, which foot goes first? when you are pushed from behind, which foot to you put forward to regain balance?
- Right
- Left
- Both

Have you ever used a foot mouse or similar foot-operated pointer?
- Yes
- No

Submit

Never submit passwords through Google Forms.

Appendix 10 - Post-study questionnaire for the study in Section 5.2

Feet Movement Times

Name

Which condition did you find the easiest to use?

- Horizontal bars with the right foot
- Horizontal bars with the left foot
- Vertical bars with the right foot
- Vertical bars with the left foot
- No difference in difficulty between the conditions

Which condition did you find the hardest to use?

- Horizontal bars with the right foot
- Horizontal bars with the left foot
- Vertical bars with the right foot
- Vertical bars with the left foot
- No difference in difficulty between the conditions

Which condition did you find the most comfortable to use?

- Horizontal bars with the right foot
- Horizontal bars with the left foot
- Vertical bars with the right foot
- Vertical bars with the left foot
- No difference in comfort between the conditions

Which condition did you find the least comfortable to use?

- Horizontal bars with the right foot
- Horizontal bars with the left foot
- Vertical bars with the right foot
- Vertical bars with the left foot
- No difference in comfort between the conditions

What did you find the hardest about using this interface?

Where was the hardest position to reach with your right foot?

Where was the hardest position to reach with your left foot?

In your opinion, what computing tasks could be improved by adding the interface you just tried?

What kind of strategy did you use when reaching the horizontal bars?

What kind of strategy did you use when reaching the vertical bars?

Submit

Never submit passwords through Google Forms.

Appendix 11 - Pre-study questionnaire for the study in Section 6.1

Demographics Questionnaire

Participant ID
To be filled in by the experimenter

Name

Age

Dominant hand

- Right
- Left
- Both

Gender

- Male
- Female
- Other:

Are you wearing vision correction?

- Glasses
- Contacts
- Need to, but not wearing
- No

How many times have you used eye tracking before?

- Never
- Once
- Twice
- Three times
- More than three times

Please rate your experience with eye trackers

	1	2	3	4	5	6	7	
No experience	○	○	○	○	○	○	○	Expert

Please rate your experience with interfaces that use non-touch gestures
(e.g. Kinect, Leap Motion, Wii)

	1	2	3	4	5	6	7	
No experience	○	○	○	○	○	○	○	Expert

Submit

Never submit passwords through Google Forms.

Appendix 12 - Questionnaire for each condition in Section 6.1.3

Translation Task

Participant ID

Interaction Technique

Please rate the selected interaction technique according to:

1 is the lowest mark
7 is the highest mark

Speed
How fast or slow was the technique for completing the tasks?

	1	2	3	4	5	6	7	
Very slow	○	○	○	○	○	○	○	Very fast

Accuracy
How accurate was the technique for completing the tasks?

	1	2	3	4	5	6	7	
Very inaccurate	○	○	○	○	○	○	○	Very accurate

Ease of learning
How easy was it to learn the technique?

	1	2	3	4	5	6	7	
Very difficult to learn	○	○	○	○	○	○	○	Very easy to learn

Ease of use
How easy was it to use the technique?

	1	2	3	4	5	6	7	
Very difficult to use	○	○	○	○	○	○	○	Very easy to use

Eye fatigue

Was the technique tiring to the eyes?

	1	2	3	4	5	6	7	
Very tiring to the eyes	○	○	○	○	○	○	○	Not tiring at all to the eyes

Hand fatigue

Was the technique tiring to the hands?

	1	2	3	4	5	6	7	
Very tiring to the hand	○	○	○	○	○	○	○	No tiring at all to the hand

Arm fatigue

Was the technique tiring to the arms?

	1	2	3	4	5	6	7	
Very tiring to the arms	○	○	○	○	○	○	○	No tiring at all to the arms

Intuitiveness

Was the technique intuitive and familiar to use?

	1	2	3	4	5	6	7	
Not intuitive	○	○	○	○	○	○	○	Very intuitive

Mental effort

How much mental work did you have to do to perform the tasks using the technique?

	1	2	3	4	5	6	7	
High mental effort	○	○	○	○	○	○	○	Low mental effort

Physical effort

How much mental work did you have to do to perform the tasks using the technique?

	1	2	3	4	5	6	7	
High physical effort	○	○	○	○	○	○	○	Low physical effort

Comfort

How comfortable was it to use the technique?

	1	2	3	4	5	6	7	
Very uncomfortable	○	○	○	○	○	○	○	Very comfortable

Suitability for the task

How suitable was the technique for the given task?

	1	2	3	4	5	6	7	
Very unsuitable	○	○	○	○	○	○	○	Very suitable

Personal Preference

How much you liked the technique?

	1	2	3	4	5	6	7	
Strongly disliked it	○	○	○	○	○	○	○	Strongly liked it

Submit

Never submit passwords through Google Forms.

Appendix 13 - Questionnaire after all conditions in Section 6.1.3

Post-study Questionnaire (Translation)

Target Translation Task

Please rank the three techniques in terms of SPEED

	1 (best)	2	3 (worst)
2D Cursor	○	○	○
3D Cursor	○	○	○
Gaze	○	○	○

Please rank the three techniques in terms of ACCURACY

	1 (best)	2	3 (worst)
2D Cursor	○	○	○
3D Cursor	○	○	○
Gaze	○	○	○

Please rank the three techniques in terms of COMFORT

	1 (best)	2	3 (worst)
2D Cursor	○	○	○
3D Cursor	○	○	○
Gaze	○	○	○

Please rank the three techniques in terms of PERSONAL PREFERENCE

	1 (best)	2	3 (worst)
2D Cursor	○	○	○
3D Cursor	○	○	○
Gaze	○	○	○

What was the best thing about your preferred technique?

Appendix 14 - Questionnaire for each condition in Section 6.1..4

Sorting Task

Participant ID

Interaction Technique

Please rate the selected interaction technique according to:

1 is the lowest mark
7 is the highest mark

Speed
How fast or slow was the technique for completing the tasks?

	1	2	3	4	5	6	7	
Very slow	○	○	○	○	○	○	○	Very fast

Accuracy
How accurate was the technique for completing the tasks?

	1	2	3	4	5	6	7	
Very inaccurate	○	○	○	○	○	○	○	Very accurate

Ease of learning
How easy was it to learn the technique?

	1	2	3	4	5	6	7	
Very difficult to learn	○	○	○	○	○	○	○	Very easy to learn

Ease of use
How easy was it to use the technique?

	1	2	3	4	5	6	7	
Very difficult to use	○	○	○	○	○	○	○	Very easy to use

Eye fatigue

Was the technique tiring to the eyes?

	1	2	3	4	5	6	7	
Very tiring to the eyes	○	○	○	○	○	○	○	Not tiring at all to the eyes

Hand fatigue

Was the technique tiring to the hands?

	1	2	3	4	5	6	7	
Very tiring to the hand	○	○	○	○	○	○	○	No tiring at all to the hand

Arm fatigue

Was the technique tiring to the arms?

	1	2	3	4	5	6	7	
Very tiring to the arms	○	○	○	○	○	○	○	No tiring at all to the arms

Intuitiveness

Was the technique intuitive and familiar to use?

	1	2	3	4	5	6	7	
Not intuitive	○	○	○	○	○	○	○	Very intuitive

Mental effort

How much mental work did you have to do to perform the tasks using the technique?

	1	2	3	4	5	6	7	
High mental effort	○	○	○	○	○	○	○	Low mental effort

Physical effort

How much mental work did you have to do to perform the tasks using the technique?

	1	2	3	4	5	6	7	
High physical effort	○	○	○	○	○	○	○	Low physical effort

Comfort

How comfortable was it to use the technique?

	1	2	3	4	5	6	7	
Very uncomfortable	○	○	○	○	○	○	○	Very comfortable

Suitability for the task

How suitable was the technique for the given task?

	1	2	3	4	5	6	7	
Very unsuitable	○	○	○	○	○	○	○	Very suitable

Personal Preference

How much you liked the technique?

	1	2	3	4	5	6	7	
Strongly disliked it	○	○	○	○	○	○	○	Strongly liked it

Submit

Never submit passwords through Google Forms.

Appendix 15 - Questionnaire after all conditions in Section 6.1.4

Post-study Questionnaire - Sorting task

Target Translation Task

Please rank the three techniques in terms of SPEED

	1 (best)	2	3 (worst)
2D Cursor	○	○	○
3D Cursor	○	○	○
Gaze	○	○	○

Please rank the three techniques in terms of ACCURACY

	1 (best)	2	3 (worst)
2D Cursor	○	○	○
3D Cursor	○	○	○
Gaze	○	○	○

Please rank the three techniques in terms of COMFORT

	1 (best)	2	3 (worst)
2D Cursor	○	○	○
3D Cursor	○	○	○
Gaze	○	○	○

Please rank the three techniques in terms of PERSONAL PREFERENCE

	1 (best)	2	3 (worst)
2D Cursor	○	○	○
3D Cursor	○	○	○
Gaze	○	○	○

What was the best thing about your preferred technique?

What was the worst thing about your least preferred technique?

« Back Submit

Never submit passwords through Google Forms.

Appendix 16 - Pre-study questionnaire for the study in Section 6.2

Personal Information

Name

Age

Occupation

Have you ever used a 3D modelling application?

- Yes
- No

Have you ever used a 3D display (e.g. watching a 3D film at the cinema)?

- Yes
- No

Do you drive?

- Yes
- No

Have you ever used a 3D mouse or similar foot-operated device?

- Yes
- No

If yes, please specify

Submit

Never submit passwords through Google Forms.

Appendix 17 - Post-Study Questionnaire for the study in Section 6.2

3D Foot Interaction

Regarding the CAMERA CONTROL task...

	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
The interaction was comfortable	○	○	○	○	○
The interaction was intuitive	○	○	○	○	○
The interaction broke the illusion of 3D	○	○	○	○	○
I felt in control of the interaction	○	○	○	○	○
I knew how to move my feet to achieve the desired goal	○	○	○	○	○
I would be able to perform an additional task with my hands	○	○	○	○	○
I felt frustrated	○	○	○	○	○
The task was mentally demanding	○	○	○	○	○

Regarding the CUBE ROTATION task...

	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
The interaction was comfortable	○	○	○	○	○
The interaction was intuitive	○	○	○	○	○
The interaction broke the illusion of 3D	○	○	○	○	○
I felt in control of the interaction	○	○	○	○	○
I knew how to move my feet to achieve the desired goal	○	○	○	○	○
I would be able to perform an additional task with my hands	○	○	○	○	○
I felt frustrated	○	○	○	○	○
The task was					

mentally demanding	○	○	○	○	○

Regarding the SPHERE SELECTION task...

	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
The interaction was comfortable	○	○	○	○	○
The interaction was intuitive	○	○	○	○	○
The interaction broke the illusion of 3D	○	○	○	○	○
I felt in control of the interaction	○	○	○	○	○
I knew how to move my feet to achieve the desired goal	○	○	○	○	○
I would be able to perform an additional task with my hands	○	○	○	○	○
I felt frustrated	○	○	○	○	○
The task was mentally demanding	○	○	○	○	○

Regarding the FOOT MENU task...

	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
The interaction was comfortable	○	○	○	○	○
The interaction was intuitive	○	○	○	○	○
The interaction broke the illusion of 3D	○	○	○	○	○
I felt in control of the interaction	○	○	○	○	○
I knew how to move my feet to achieve the desired goal	○	○	○	○	○
I would be able to perform an additional task with my hands	○	○	○	○	○
I felt frustrated	○	○	○	○	○
The task was mentally demanding	○	○	○	○	○

In your opinion, what are the main advantages of this interface for these interactions?

And what are the main disadvantages?

What was the most difficult aspect of these interactions?

Submit

Never submit passwords through Google Forms.