Perception of Body and Hand Animations for Realistic Virtual Characters

Sophie Jörg

Dissertation

Presented to the

University of Dublin, Trinity College

in fulfilment

of the requirements

for the Degree of

Doctor of Philosophy

University of Dublin, Trinity College

March 2011

Declaration

I, the undersigned, declare that this work has not previously been submitted as an exercise for a degree at this or any other university and that it is my own work unless otherwise stated.

I, the undersigned, agree that Trinity College Library may lend or copy this thesis upon request.

Sophie Jörg

March 7, 2011

Abstract

Researchers and artists bring astonishing cartoon style characters and fantasy creatures to life. However, despite great improvements in recent years, creating a persuasive and highly realistic virtual human is still a big challenge. The motions of a virtual character in a movie or in a game have a large impact on the observer's impression of it. People are remarkably good in recognizing human motion and very subtle details can be distinguished and interpreted differently. Every person moves and walks in her own way, but all these different styles are perceived as human motions. However, minor changes can lead observers to classify the motion as unnatural. Correct motions are therefore crucial in order to achieve compelling characters.

Motion capture technology is commonly used to bring highly realistic characters to life. Even though this technology generates impressive results it is not without flaws and can produce motion artefacts. The impact of such errors is not well understood so far. Conventional perceptual experiments with short snippets of animation are valuable for measuring straightforward perceptibility thresholds such as just noticeable difference. However, this approach does not capture the higher level cognitive responses of complex scenarios, such as those found in movies or games.

We therefore introduce and apply a method to study the effect of virtual characters on the viewer that evaluates changes in the emotional response, attention to details, interpretation, and perceived quality of an animation due to degradations in the motion. As the results of our experiments highlighted the importance of correct finger motions, we then focus on the perception, understanding, and enhancement of finger animations.

To generate highly realistic animations, we use optical motion capture technology. The accurate capturing of finger motions is still challenging and error-prone due to the small scale and dexterity of fingers. Thus, we improve existing techniques to reach the required accuracy for the creation of our stimuli. Our perceptual approach uses *vignettes*, i.e. longer animations that tell a story. Within three experiments, where we investigate people's reactions to synchronization errors and degradations in human motion with questionnaires and by tracking their gaze, we steadily revise and improve this method. To get further insight into finger motions, we perform two perceptual studies, which analyse the noticeability of synchronization errors and the perceived quality of four types of hand animations. Additionally, we examine and reduce the dimensionality of finger motion by identifying barely noticeable movements and rotations that can be accurately computed as a linear relation of each other. Finally, we present a pipeline to add finger motions to a crowd animation system as a platform for further experiments.

Acknowledgements

First of all, I would like to thank my advisor, Carol O'Sullivan. I sincerely appreciate all of the opportunities, guidance, and enthusiasm she has given me and shown over the past years. In my first year, she introduced me to Jessica Hodgins, who gave me the opportunity to visit Carnegie Mellon University and Disney Research, Pittsburgh. I am very grateful for this experience and thank Jessica for all her advice and support.

Next, I would like to thank my officemates and colleagues at GV2, Trinity College Dublin, and the Graphics Lab, Carnegie Mellon University, for their support and feedback. Thanks to Catherine Pelachaud and Gerry Lacey for being my examiners and making the viva so enjoyable.

Capturing the motions of multiple actors is not a solitary task. I would like to thank all of the people who helped me creating animations, especially Moshe Mahler for creating models and spending endless hours changing and rendering animations, Justin Macey for providing assistance for motion capturing and labelling, Sang Il Park for creating facial animations, and all the others who spent an hour or two doing the capture sessions. Special thanks to all of the colleagues and friends who agreed to wear a tight Lycra suit for research, namely Fintan, Melissa, Kerstin, Serena, Cathy, and Micheál, and the actors, Tami, Jeff, and Daniel, for their great performances.

Thanks to all of the people who helped me carry out experiments: my co-author Rachel McDonnell, with whom I worked in the first perceptual projects of my research; the colleagues and friends who gave their feedback for pilot experiments, especially Tom, Fintan, Daniel, and Martin; everyone who helped recruit new participants or carry out the experiments, including Marina and Cathy; the lecturers who gave me part of their lecture times; the independent investigators who evaluated questionnaires; and of course many thanks to all of the participants who volunteered to take part in the experiments!

This research has been enabled by the Embark Postgraduate Scholarship Scheme of the Irish Research Council for Science, Engineering, and Technology (IRCSET) and furthermore supported by Science Foundation Ireland within the Metropolis Project and Disney Research, Pittsburgh.

Finally, I would like to thank my fiancé Alex, my family, and my friends for supporting me during the past years, giving me shelter between my rather frequent moves, and cheering me up when needed.

Relevant Publications

Journals

1. **The saliency of anomalies in animated human characters**, Jessica Hodgins, Sophie Jörg, Carol O'Sullivan, Sang Il Park, and Moshe Mahler, ACM Transactions on Applied Perception (TAP), July 2010, volume 7, issue 4, article no. 22.

2. **Investigating the role of body shape on the perception of emotion**, Rachel McDonnell, Sophie Jörg, Joanna McHugh, Fiona N. Newell, and Carol O'Sullivan, ACM Transactions on Applied Perception (TAP), August 2009, volume 6, issue 3, article no. 14.

3. **Evaluating the effect of motion and body shape on the perceived sex of virtual characters**, Rachel McDonnell, Sophie Jörg, Jessica Hodgins, Fiona Newell, and Carol O'Sullivan, ACM Transactions on Applied Perception (TAP), January 2009, volume 5, issue 4, article no. 20.

Papers

1. **The perception of finger motions**, Sophie Jörg, Jessica Hodgins, and Carol O'Sullivan, Proceedings of the 7th Symposium on Applied Perception in Graphics and Visualization, July 2010, pp. 129–133.

2. **Exploring the dimensionality of finger motion**, Sophie Jörg and Carol O'Sullivan, Proceedings of the 9th Eurographics Ireland Workshop (EGIE), December 2009, volume 9, pp. 95–101.

3. **Evaluating the emotional content of human motions on real and virtual characters**, Rachel McDonnell, Sophie Jörg, Joanna McHugh, Fiona Newell,

and Carol O'Sullivan, Proceedings of the 5th Symposium on Applied Perception in Graphics and Visualization (APGV), August 2008, pp. 67–74.

4. **Virtual shapers & movers: form and motion affect sex perception**, Rachel McDonnell, Sophie Jörg, Jessica K. Hodgins, Fiona Newell, and Carol O'Sullivan, Proceedings of the 4th Symposium on Applied Perception in Graphics and Visualization (APGV), July 2007, pp. 7–10.

Posters

1. **What's the effect of degraded human motion?**, Sophie Jörg (supervisors: Carol O'Sullivan (TCD) and Jessica Hodgins (CMU)), GV2 workshop, May 2009.
2. **Perception of errors in synchronization**, Sophie Jörg (supervisors: Carol O'Sullivan (TCD) and Jessica Hodgins (CMU)), Irish Graduate Student Symposium on Vision, Graphics and Visualisation (VGV08), June 2008.
3. **Perception of synchronisation errors in hand motions**, Sophie Jörg, Jessica Hodgins and Carol O'Sullivan, ACM SIGGRAPH/Eurographics Symposium on Computer Animation (SCA'08), July 2008.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Character animation is one of the most compelling and challenging fields in computer graphics. Virtual characters entertain and engage us in movies and games and they inform or train us in virtual reality applications related to areas as varied as medicine, education or military training. Tremendous progress has been made in the creation of realistic human-like virtual characters in recent years. Images of virtual humans can be of such high quality that one cannot say with full certainty if they depict a real person or not. Nevertheless, as soon as such characters move, we can almost instantly recognize that they are artificial. A character animation Turing test has not yet been passed for moving human-like realistic characters. One reason for this is that humans possess impressive abilities to recognize and interpret human motion.

The aim of our research is to increase the perceived quality of the motions of anthropomorphic characters. To further improve the creation of compelling virtual characters we need to find ways to analyse and understand the impression that they make on us. The questions we ask include: How much can natural motion be altered before an observer notices it? Are some kinds of errors in motion less important than others? What are the main points an animator has to keep in mind to create believable, lifelike virtual humans? We present an approach using *vignettes*, i.e. short animated movies that convey emotional content, to evaluate the perceptual impact of various types of errors and anomalies in motion. Building on previous work in computer graphics, we do not only test the perceptibility of an error, but also study changes in the emotional content, attention to details, interpretation and perceived quality of the scene. We present a set of perceptual experiments to illustrate our approach. Our findings demonstrate that hand motions play a role

in conveying emotions and can also change the interpretation of a scene. We therefore study hand motions in a subsequent set of experiments. Finally, we present a method to reduce the dimensionality of finger motions to facilitate better capture and animation of finger motions and we show how to add hand motions to the characters of a crowd animation system.

1.1 Motivation

Researchers and artists bring astonishing virtual characters to life with computer graphics. However, most major movies use highly stylized characters, animals, or fantasy creatures to narrate their stories as can be seen in *Ratatouille* (2007) and *Up* (2009), both produced by Pixar/Walt Disney Pictures, *Shrek Forever After* (2010), produced by Dreamworks, *Igor* (2008) created in Sparx Animation Studios, or *Ice Age: Dawn of the Dinosaurs* (2009) by Blue Sky Studios to mention only a few recent examples. In other movies the characters interact with real humans, but are themselves not fully human-like, such as Gollum in *The Lord of the Rings: The Two Towers* (2002) or the Na'Vi in *Avatar* (2009) (see Figure 1.1 (a)). The first movie, which seriously tried to depict realistic virtual humans was *Final Fantasy: The Spirits Within* (2001), animated by Square Studios. The movie got very mixed reviews. It was recognized as a technical milestone, but was highly criticized for its plot [Rot11a] and failed at the box-offices [Fil11]. The next movie seeking realism was *The Polar Express* (2004) by Robert Zemeckis, which was widely criticized for the lack of life in its characters, who were described as being "creepy" or "cold" [Rot11b] (see Figure 1.1 (b)). Impressive advances in terms of realism have been achieved with *Beowulf* (2007) also directed by Robert Zemeckis, but there are still many improvements to be made to create a persuasive and highly realistic virtual human.

When viewing sequences with very human-like virtual characters, audiences tend to feel a sense of eeriness or revulsion rather than empathizing with the characters or becoming emotionally engaged in the story. Numerous articles in the public press and elsewhere [Gel08] have speculated that these more realistic characters occasionally fall into the uncanny valley hypothesized by Mori in 1970 [Mor70]. According to Mori, the familiarity of a character grows with increasing human likeness up to a certain point, where the familiarity drops for characters that are close to humans, but not quite there, until it reaches its highest point for real humans. The concept is far from precise and it is not clear what

(a) (b)

Figure 1.1: (a) Picture from the movie *Avatar* (2009). The characters received good audience reactions, but are not human, ©20th Century Fox. (b) Picture from the movie *The Polar Express* (2004). The human-like characters were criticized as being "creepy" or "cold", ©Warner Bros. Pictures.

the axes of the graph should be. Indeed, it is unlikely that viewers' complex, multisensory responses to virtual characters can be described by such a simple graph [TG09a]. Nevertheless, it provides a useful conceptual framework within which to explore these important issues. Many different causes for this supposed revulsion have been put forward, including motion artefacts, incorrect eye gaze, stiff faces, or the appearance of diseases. Given our limited understanding of this phenomenon, we need techniques to analyse the impact of the improvements and remaining anomalies in human motion and appearance on the perception and interpretation of a scene. We need to not only analyse perceptibility thresholds, but to evaluate higher-level cognitive and emotional responses.

Mori hypothesized a stronger effect when motion is involved. We know since Johansson's experiments [Joh73, Joh76], that humans have astonishing abilities to recognize and interpret very subtle changes in motion. Johansson developed a technique to separate the investigation of human motion from shape or visual appearance by attaching reflective dots – point-lights – to various parts of the body and recording an actor so that only the points and not the shape of the body can be seen [Joh73]. When people viewed the point-light pattern of a walking motion for an exposure time of only 0.2 seconds, all of them effortlessly recognized a walking person. With an exposure time as short as 0.1 seconds, still 40% of participants reported a walking person [Joh76]. A still frame of the same points, however, was never interpreted as a human body within his experiments [Joh73]. Later research underlined how details in human motion can be distinguished and used to draw conclusions, for example to recognize the sex of a walker [CK77], specific people [KC77], or emotions [ADGY04]. These results reinforce our intention to focus our work on the perception of motions. Even though motion capture technology nowadays enables us to measure and record human motions very accurately, the technology is not without

flaws and we specifically take into account errors that happen when using motion capture.

We began our research with an experimental investigation of the importance of synchronization in human animation. As the results of these experiments pointed out that the hands and fingers play a role in the interpretation of a scenario and when conveying emotions, we study this aspect in more detail. Human hand motions are omnipresent. We use our hands to communicate ideas and to perform complex tasks. Motion capturing hand movements differs from the capturing of body motions because of the high number of small bones involved, which allow only a limited number of small markers to be used. When more accuracy is needed, the cameras of an optical motion capture system need to be closer to the markers, which is likely to reduce the capture space. Post-processing of hand motion data is characterized by labelling errors and occlusions resulting in time-consuming manual labour. Therefore, in most capturing sessions for games and movies, only the body movements are captured whereas the hands are animated manually. In this work, we therefore evaluate different methods to animate hands and make recommendations to improve the animation of fingers and hands. Although, strictly speaking, *finger* denotes four digits and excludes the thumb [PFS06], for simplicity we use this term to refer to all five digits.

1.2 Objective and Contributions

The overall objective of our research is to contribute to the creation of more compelling virtual anthropomorphic characters. To achieve this goal, we first carry out perceptual studies to try to understand how people recognize and interpret human motion. Our aim is to capture the higher-level cognitive and emotional responses that the complex scenarios found in movies or games are intended to elicit. As our findings underline the importance of hand motions on the viewer, we then study the motions of hands in greater detail and develop techniques to improve the animation of fingers.

Our contributions are as follows:

- Improvements to existing motion capture technology to capture details in human motion. In particular, the accurate capturing of finger motions is still challenging and error-prone due to the small scale and the dexterity of fingers, so we enhance current optical motion capture techniques. Furthermore, we capture facial motions

and gaze behaviours and integrate them into our animations.

- A method to analyse the perception of complex animations based on the higher-level cognitive and emotional responses of the viewer. Our approach evaluates the influence of different aspects in animation on the emotional content, attention to details, interpretation, and perceived quality using *vignettes*, i.e. short animated movies that convey emotional content. We use, improve, and validate our approach within three studies, which measure reactions through questionnaires and eye tracking of participants viewing high quality motion captured scenes. We first investigate the perceptual consequences of errors in synchronization caused by combining separately captured human motions, e.g. hands with full body, upper with lower body, or two characters interacting. Our scenarios include an angry reaction to a computer crash, an argument between a couple, and an impatient woman waiting for a bus. We find that errors in synchronization lead to changes in the meaning or affective content of an animation, particularly when physical interactions are present. Furthermore, we show that erroneous finger animations can change the interpretation of a scene even without altering its perceived quality. In our second experiment, we study the impact of degradations on the viewer. We developed three vignettes picturing angry scenarios about an unhappy relationship and altered them by leaving out or simplifying parts of the animation. For example in one condition the facial animation was discarded, resulting in an unnaturally stiff face. We then investigated the reactions of viewers and found that changing different aspects of multisensory content can lead to interaction effects, with the sound having a high influence. Our third study focuses on changes in human motions, some of which resemble certain disease symptoms. Our results show that facial anomalies are more salient than body motion errors.

- New results on the perception of hand motions. Specifically, we assess which one out of four hand animation techniques is best and how much synchronization error can be perceived in hand motion. In two experiments, designed to investigate finger animations, we asked the following questions: When are errors in finger motion noticeable? and: What animation method should we recommend? We found that synchronization errors of as little as 0.1s can be detected, but that the perceptibility of errors is highly dependent on the type of motion. In our second study, out of the four conditions tested – original motion capture, no motions, keyframed animation

and randomly selected motions – the original motion captured movements were rated as having the highest quality.

- Reductions in the recorded degrees of freedom of hand motions. We analyse pairwise correlations and the ranges of joint orientations. Thereby, we find out how we can reduce the dimensionality of finger motions while keeping individual joint rotations as a basis for the motions, thus simplifying the task of capturing and animating hand motions.

- A pipeline to enhance a crowd animation system with finger motions. With this pipeline, a system depicting a virtual city can be used as a testbed for further investigations in the perception and creation of finger movements.

The results of this research will be useful for the game and animation industries. Furthermore, they will contribute to our general understanding of human motion and can therefore have an impact in fields such as robotics, psychology, or neuroscience.

1.3 Overview of Chapters

The remainder of this thesis is organized as follows:

Chapter 2 reviews previous research related to our work. We provide an overview of methods to create and modify motion captured animations, including dimensionality reduction. Previous studies on the perception of human motions and virtual characters are also detailed.

Chapter 3 introduces our methodology. We first describe how we created our stimuli, i.e. how we used and improved motion capture technology to capture finger and facial motions and how we captured the gaze behaviours of our actors. We then describe the research methods on which our experiments are based by giving an overview of data acquisition methods, such as questionnaires or eye-tracking and detailing our data analysis methods.

In **Chapter 4** we present our evaluation method to analyse errors in the motion of animated humans using vignettes. We illustrate our approach with three studies that investigate the effects of errors in synchronization, the impact of degrading different aspects of animation and the saliency of anomalies in animated human characters.

Chapter 5 contains our work on hand movements. We detail two studies on the perception

of finger motions, present our approach to reduce their dimensionality, and describe a pipeline to include finger animation into a crowd animation system.

Chapter 6 summarizes our findings and presents future areas of work that merit further investigations.

Chapter 2

Related Work

The techniques and algorithms to create virtual characters have been improved at a very quick pace in recent years. The first completely computer animated feature film, *Toy Story*, by Pixar Animation Studios, was released just 15 years ago in 1995. Nowadays, several CG animated movies and multiple games featuring virtual 3D characters are released every year, virtual humans interact and train people in virtual reality applications, and virtual characters even share the screen with real humans. To further advance the creation of compelling animations, the perception of motion is increasingly taken into account and the influence of different factors and algorithms in the animation of virtual characters is investigated. This chapter describes prior research related to the animation and perception of virtual characters.

First, we provide an overview of the methods to animate realistic human-like characters. We briefly introduce general animation concepts before we discuss, in more detail, motion capture techniques as these are most commonly used for realistic anthropomorphic characters. We then focus on methods to generate sophisticated hand animations. This research is related to our development of a motion capture pipeline for detailed hand capture presented in Chapter 3 and to the creation of all the animations used within this thesis. The section is completed by presenting methods to reduce the dimensionality of motion databases, which is relevant to our approach to analyse the dimensionality of finger motion in Section 5.2.

Second, we discuss research on the perception of motion. The recognition of human movements independently of their shapes was first studied in psychology. In computer graphics, research on the perception of virtual characters aims at a better understanding

of the factors that influence our impressions. We then describe one of those factors: the influence of the geometric model as it is relevant to our study on degradations in animation presented in Section 4.2. Finally, we discuss the assumption of the *uncanny valley*, which was our inspiration for the perceptual studies presented in Chapter 4. The study of the perception of motions for real and virtual characters is not only useful for the animation of characters in movies and games, but is also relevant in fields such as biomechanics, psychology, neuroscience, robotics, ergonomics, computer vision, and virtual reality, which is why our discussion presents research from several of these areas.

2.1 Motion Generation

In 3D animation, the body of a virtual character is represented by a large number of vertices (data structures that describe a point in space), which are assembled into polygons, or by a large number of control points defining cubic patches, usually NURBS [Par07]. These polygons or patches form a mesh. To animate such a mesh, the position of every point has to be defined for every frame. An established technique to simplify this procedure is the use of a hierarchical, rigid-body structure, which reproduces a simplified version of a human skeleton (see Figure 2.1). Each point of the mesh is assigned to one or several bones or joints, a process called *weighting* or *skinning*. The pose of a virtual character can now be fully defined by specifying the position and orientation of the upper-most joint in the hierarchy, called the *root*, and the orientation of every further joint. The majority of virtual characters are based on skeletons, although more sophisticated control techniques than skinning can be programmed. For example, muscle bulging can be automatized to happen when the limbs of a character are moved, or controls at the fingertips can be used to modify the posture of the whole hand. The method to set up a character and to generate high-level controls for multiple degrees of freedom to simplify the task of animation is referred to as *rigging*.

2.1.1 Animation and Simulation Methods

An animation consists of a set of coordinates defining the position and orientation of the hierarchical structure or skeleton joints for every frame. To specify each coordinate for each single frame would be a very tedious task. Various techniques have been developed to deploy the skills of an animator more effectively. *Keyframes* – the coordinates for a

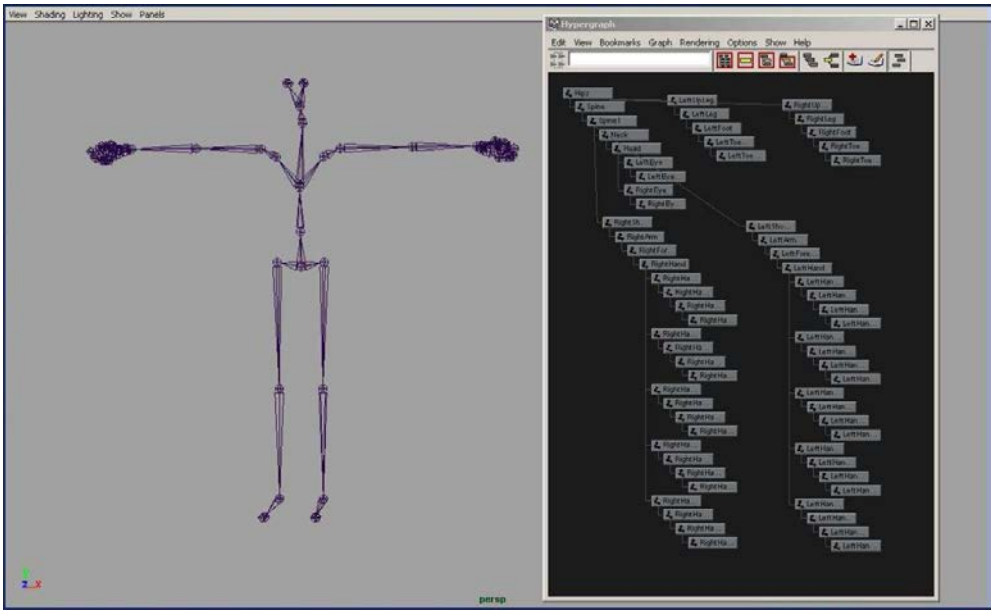


Figure 2.1: Typical skeleton (left) and hierarchy (right) for a human-like virtual character. The great number of small joints of the hands can be seen very well.

specific frame – are defined for selected frames between which the motion is interpolated automatically. Two basic kinematic approaches to position the skeleton are known: *forward kinematics*, in which the joints are animated from the root to the tips, and *inverse kinematics*, where only the target coordinates for an end effector, such as the hand to move the arm, are set and the joint orientations of the hierarchical structure are solved automatically [Par07]. Additionally, constraints can be defined to set limits to joint orientations or positions, for example to avoid collisions.

To enhance the realism of an animation, physically based simulations can be used, which are based on real-world parameters such as the estimated mass and the calculation of forces and energies. These approaches have the advantage of being physically correct, which is a first step in being realistic. Although great results have been achieved to synthesize inanimate materials such as fluids or snow, a human is very complex to simulate and many parameters need to be estimated. Hodgins et al. [HWBO95] used a dynamic simulation combined with control algorithms to create three types of motions: running, bicycling, and vaulting. However, this type of approach necessitates new control algorithms for each type of motion, making the development of new motions very time-consuming.

In the early 1990's, a new technique became increasingly common: motion capture. We use motion capture in this dissertation to create our stimuli. Furthermore, we refine existing motion capture techniques to capture finger motions as accurately as possible. An

introduction to motion capture technology and its applications is therefore presented in the next section. An overview of further algorithms and techniques in computer animation can be found in computer animation books, such as [Par07].

2.1.2 Motion Capturing

Motion capture is "the technology that enables the process of translating a live performance into a digital performance" [Men99]. Motion capture systems have been used commercially since the 1990's. Numerous movies, especially those that aimed for photorealistic characters, used motion capture. Examples are *The Polar Express* (2004), *Beowulf* (2007) and *A Christmas Carol* (2009) all directed by Robert Zemeckis, *Final Fantasy: The Spirits Within* (2001) (see Figure 2.2 (a)), the character of *Gollum* in *The Lord of the Rings: The Two Towers* (2002) and *The Lord of the Rings: The Return of the King* (2003), or the Na'vi characters in *Avatar* (2009). Nearly all games with photorealistic human-like 3D characters nowadays use motion capture, such as the *NBA* video games (1989-2010) (see Figure 2.2 (b)), the *FIFA* video games (1993-2010), *Fallout3* (2008), or *Heavy Rain* (2010) to name just a few examples.

(a) (b)

Figure 2.2: Examples of virtual characters animated with motion capture: (a) promotional shot of the character Aki from Final Fantasy, ©Square Co., Ltd; (b) screen shot from NBA2K11, ©2K Sports.

Available motion capture techniques include the following: optical with active or passive markers, optical without markers, magnetic, acoustic, inertial, and electromechanical as well as specific developments for hands and face. In general, a varied number of sensors or markers are attached to specific points of the body. Depending on the technique used, the sensors can measure their position or orientation or both or their flexion. Based on this data, a virtual skeleton is calibrated and animated, see [HFP*00, HSD05, KOF05], and [WWXZ06] for optical motion capture or [BRRP97], and [OBBH00] for magnetic

systems. Overviews of the different available techniques can be found, among others, in [Men99, Par07, KW08].

As motion capture records real motion, it produces very lifelike movements. Its usage in movies and games therefore depends on the creative goal of the project. Cartoon effects, such as an exaggerated squash and stretch – one of the 12 principles of animation [TJ81, Wil01] – are usually animated with different techniques. Our aim is to produce very realistic human-like stimuli. Therefore, motion capture is our method of choice.

Optical marker-based systems allow the highest accuracy. Current state-of-the-art systems use a set of 40 to 50 individual markers to capture the whole body (see Figure 2.3), although there have also been studies with over 300 markers [PH06]. The rise in popularity of motion capture has lead to the development of more sophisticated systems and algorithms. Still, motion capturing is far from being perfect and errors and artefacts occur during the capturing process. One of the topics of this thesis is how these artefacts affect the viewer's perception of motions.

Figure 2.3: Actor in a motion capture suit being tracked with the GV2 Vicon optical motion capture system.

The variety of possible human motions is too large to be captured, which makes it necessary to develop algorithms and techniques to edit motion captured motions while still producing natural looking results. We give a brief overview of relevant research from this area in the next section.

2.1.3 Modifying Motion Captured Data

A major challenge of animating virtual characters with motion capture is the difficulty of adapting the motions to be able to reuse them in different situations. Slight changes to increase the variability, as can be easily produced with simulated animations, cannot be generated directly. In order to use captured motions in a game or for new scenes, clips of motion have to be rearranged and combined in a reasonable manner. Furthermore, it must be possible to alter existing motions to meet specific constraints. Many algorithms have been proposed to edit and expand motions with the goal of producing natural looking results.

Kovar et al. [KGP02] and Lee at al. [LCR*02] present methods to automatically assemble new motions. In [KGP02], appropriate transition points are chosen based on a distance metric that takes the position of the mesh of the virtual character into account. A so-called *Motion Graph* is generated in which the edges of the graph contain clips of the original motion data or transitions. At each node of the graph the next edge/motion clip can be selected. By searching paths through that graph, new motions are obtained, which might fulfill specified goals.

By warping the motion curves of an animation to fit a small set of keyframes, Witkin and Popovic [WP95] introduced a simple algorithm (not restricted to motion capture) to alter and reuse motions in a way that makes them satisfy constraints. Bruderlin and Williams [BW95] present a variety of methods from image and signal processing, such as multiresolution filtering or motion displacement mapping, to alter motions. To avoid the repetitiveness of motions and to add more variability to motions, Bodenheimer et al. [BSH99] attempt to add noise. This is exemplified by adding a continuous noise function to the sensors of the simulation. A short perceptual analysis presented by the authors does not give a clear answer as to the level of noise that would be appropriate. A drawback of these techniques is that there is no guarantee that the resulting motion will look natural. Big changes tend to look distorted and unnatural. Safonova and Hodgins [SH05] analyse the physical correctness of interpolated human motion and suggest new methods to keep interpolated motions physically correct. But even physically valid motion is no guarantee of looking realistic as resulting animations might also evoke rag-doll motions.

Most algorithms that are recommended to alter motion captured movements are only superficially evaluated, if at all, even though they might cause considerable artefacts. It

is very difficult to predict which algorithms might cause believable motions. Ikemoto and Forsyth [IF04] present another study in which various parts of motion are assembled. To recycle motions and expand a database, they cut limbs from one motion sequence and attach them to another. As expected, this does not always result in correct human motions. Instead of avoiding non-human-like motions, their approach uses a motion classifier, that subsequently attempts to sort human and non-human movements. Even though a classifier can be trained to recognize self-interpenetrations or fast twitches, this ability is not enough to judge the quality of resulting motions. Furthermore, the presented classifier takes only single frames into account, thereby ignoring any type of time based motion errors. We hence need better tools to evaluate alterations of motion captured movements.

Majkowska et al. [MZF06] suggest capturing hand and body motions separately and demonstrate how to synchronize the motions automatically using a dynamic time warping algorithm. This technique requires the performer to repeat the acted motion very precisely, and it is possible that artefacts may occur. Here as well, a more profound analysis of the resulting motions is missing. These approaches to motion edits were our motivation to analyse specific types of errors in a perceptual study, i.e. synchronization errors between hand and body motions and lower body vs. upper body (see Section 4.1).

2.1.4 Creating Hand Motions and Gestures

Capturing the subtleties of finger motions is especially challenging, simply because hands and fingers are much smaller than the entire body, and because of the relatively high number of degrees of freedom found in hand movements. Frequent occlusions increase the complexity of the task. In Section 3.1, we describe our improvements to the motion capture pipeline to capture hand and body motions for multiple characters simultaneously. Kitagawa and Windsor [KW08] describe the importance of capturing hand motions to convey emotions but also mention the challenges. They list five possible ways, in which hand motions can be captured with an optical motion capture system in a production setting. The most common method is probably to only capture the wrist orientation with three markers on the back of the hand. By adding one marker on one fingertip (four markers in total), basic bending of that finger is captured and plotted to all finger joints. One extra marker on the thumb tip (five markers in total) enables a separate measurement of the thumb's motions. One further marker on a different fingertip (e.g. to have thumb, index and pinky) adds more motion possibilities. Finally, a complete capture of the hand

can be achieved with a suggested 21 markers. Nevertheless, a complete capture of the hand is rarely recorded due to the described occlusions and differences in scale, making time-consuming post-production necessary.

Given those constraints, hand motions are often animated manually or captured with glove-based techniques in productions. Among the latter, the CyberGlove is currently the most professional dataglove commercially available [Cyb11] (see Figure 2.4). A set of 18-22 bending sensors measure the pose of the hand. General drawbacks of gloves are their poor accuracy and their time-consuming and complex calibration [KHW95, KZK04]. Other examples of commercially available datagloves, such as the 5DT Data Glove [Fif11] or the DG5 VHand [DGT11], only use five bending sensors – one per finger – to measure the posture of the hand, which does not fulfil the accuracy that we aim for in our stimuli. A wide variety of glove based systems has been suggested. An overview can be found in the surveys of glove-based systems presented by Sturman and Zeltzer [SZ94] and Dipietro et al. [DSD08] or on the webpages of commercial resellers, such as Virtual Realities [Vir11].

Figure 2.4: Pair of CyberGloves made by CyberGlove Systems.

Progress has been achieved in computer vision algorithms to recognize hand postures based on one or multiple cameras [AS03, EBN*07, PSH97, WP09]. However, the goal of these techniques is usually to identify a hand posture out of a set of possibilities in a controlled environment, when only the hand or a small environment is present, so that hands can be used as user interfaces. They are therefore not suitable for our purposes as we are looking to capture subtleties of hand movements accurately.

Once motions have been captured, techniques exist to alter them. For example, Pollard and Zordan [PZ05] introduce a method which combines physically based animation and hand motions from a database, while Kry and Pai [KP06] present their approach which modifies motion capture data when contact is involved. In addition to the hand motions

they capture the contact forces to extract joint compliances. The captured motions can then be retargeted by adapting the joint compliances.

The approaches to animate the body of a virtual character are used to generate motions for the hands as well. Approaches, using inverse kinematics, where only the coordinates of an end effector are defined and the underlying structure is computed have been described, e.g. by El-Sawah et al. [ESGP06] and Aydin and Nakajima [AN99]. Albrecht et al. [AHS03] animate an anatomically correct hand model using muscle contraction values given over time. Even though the results look realistic, estimating the correct muscle contraction values to obtain a desired result is time-intensive and not straightforward. Sueda et al. [SKP08] developed their method to perform the other way around. Based on an existing animation, they generate the movements of the tendons and muscles under the skin to create a more realistic representation.

Algorithms to compute grasps have been studied extensively in the field of robotics [BK00], but are also well represented in computer animation. Liu [Liu08, Liu09] uses physics-based methods to synthesize hand manipulation. Li et al. [LFP07] present a shape matching algorithm that matches hand shape to object shape to choose a close grasp from a database.

Another area of research where hands of virtual characters are studied is research on conversational agents. One goal in that area is to automatically generate a virtual character, which speaks with a correct intonation and moves its head, eyes, and hands automatically in a believable way, only based on written text. Cassell et al. [CVB01] present the Behavior Expression Animation Toolkit (BEAT), which translates text into synthesized speech and synchronized nonverbal behaviour (body gestures and facial animation). Their approach is based on a set of rules derived from nonverbal conversational behaviour research. To create body gestures the system basically evaluates when to use specific types of gestures, for example, gestures representing contrast or gestures corresponding to particular actions known by the system. When none of the other gestures is used and it is appropriate, it performs beat gestures. Interestingly, the system allows to define hand shapes and arm trajectories independently to create a gesture.

Conversational agents show the importance of believable gestures, as incorrect or badly synchronized motions are salient. Furthermore, they provide an ideal testbed in which to investigate specific features, such as the perception of expressive behaviours [Pel05] or the impact of finger movements.

2.1.5 Dimensionality Reduction

Motion capture enables the creation of large databases of realistic motions. The amount of data in character animation has increased massively since the use of motion capture technology has become more widespread. In Section 5.2, we present our approach to reduce the dimensionality of hand motions and thereby to simplify the capture, animation, and synthesis of hand motions. We compute the best subset of our current basis by reducing motion curves with a small range and by finding curves that can be well approximated through a linear combination of other curves.

A multitude of techniques has been developed and adapted to reduce the size of data or to structure and analyse it. Principal Component Analysis (PCA) is probably the most common method to compress high-dimensional data [SFS98, CGA07, BZ04]. PCA finds a new orthogonal basis that represents the data, so that the greatest variance is represented in the first coordinate, and each further coordinate (also called principal component) accounts for the next greatest variance. The data can then be represented with a lower number of dimensions, such that the error is as small as possible, by retaining only the first, largest components and disregarding the rest. PCA can vastly reduce the dimensionality of motion data while only producing small errors. Safonova et al. [SHP04] showed that full body motions performing specific behaviours can be reduced from the 60 DOF to a space with 6 to 10 basis vectors using PCA. To fix errors, such as foot slipping, that happen with such a low-dimensional basis, they solve an optimizer function that keeps the motion close to the low-dimensional representation while user-specified criteria and physics constraints are satisfied.

Arikan [Ari06] presents a method to compress motion capture databases using Bézier curves and clustered PCA. A JPEG like compression algorithm is used for the feet to preserve contacts. Amongst other results, they show that the root mean squared error (RMS) is unsuitable to measure the visual fidelity of a compressed motion, as different motions with the same RMS error can yield very distinct motion qualities, especially when high frequencies or contacts are involved.

Studies that are particularly concerned with hand motions can be found in the fields of biology, neuroscience, robotics and computer vision. The analysis and development of grasping techniques is an extensive subject of study. Santello et al. [SFS98] show that the movements of 15 measured degrees of freedom of the hand performing 57 different grasps

could by represented to more than 80% accuracy by the first two components of a PCA. Braido and Zhang [BZ04] find that as much as 98% of the variance can be explained by the first two components of PCA, though only four fingers and two types of gestures were analysed. Ciocarlie et al. [CGA07] use PCA and *eigengrasps* to plan grasp motions for robotic hands with highly reduced dimensionality. Nevertheless, one disadvantage of PCA is that the resulting components can be hard to interpret. As the basis completely changes, a specific finger joint would be affected by several components, which is not very intuitive for an animator. For example, Santello et al. [SFS98] find that changing the value of their first component, having set the coefficients of all other principal components to zero, varies the hand shape from a spider-like shape, with large abduction for the index and a large adduction for the pinky, to a more closed shape, with the ring and little finger bent further than the index and middle finger. It would be very difficult to generate a specific hand shape or to animate a hand manually by using these components. Furthermore, from specific values for the components it is difficult to know how the fingers would be oriented. An animation method that relies on the standard joint orientation axis of the finger joints is far easier to understand and to control for an animator.

Further methods have been developed to reduce the dimensionality of hand motions. Rijpkema and Girard [RG91] use a knowledge-based approach to reduce the dimensionality of hands to develop a grasping strategy. As an example, they approximate the relationship of the flexion of the proximal interphalangeal joint (PIP) with the flexion of the distal interphalangeal joint to

$$f_{DIP} = \frac{2}{3} * f_{PIP} \tag{2.1}$$

based on several measurings of human subjects. Of course this type of approximation would be too simple when a higher accuracy is needed as in our work, but for basic animations it can be very useful. Chang et al. [CPMX07] show that a classifier for hand grasps retains 92% of its accuracy when trained with five optical motion capture markers compared to a classifier trained with a full marker set of thirty markers, which indicates again that many hand motions can be described with few degrees of freedom.

In addition to specific methods to reduce the dimensionality of hands, a variety of metrics have been developed to measure the similarity of curves in general. Examples are correlation coefficients, time warping techniques, the search for the longest common subsequence [VGK02], or the extraction and comparison of features [OFH08]. In this thesis, the root mean squared deviation is used as an initial metric in our approach described in Section

5.2, but further types of curve similarity metrics might be promising.

Although the research on motion capture and editing techniques shows an impressive variety of possibilities, many details remain unresolved. Furthermore, many of the presented algorithms do not guarantee human-like motion. In this thesis we examine how errors in animation, such as those resulting from changes to motion capture data, alter the perception of motion of virtual characters. To this end, we present studies on the perception of real and virtual human motion in the following section.

2.2 Motion Perception

When we create a movie or game, we would like to be able to assess and to measure the effect of our work. How compelling are the characters that we created? How could we make them more convincing? Where can we save computing time? Starting with the general recognition of human movement, we review research on the perception of motion for virtual characters and on the influence of the geometric model on motion perception. Finally, we present the assumption of the *uncanny valley*, which is a motivation for some of our perceptual experiments.

2.2.1 Perception of Human Motion

People are remarkably skilled at recognizing human motion. Johansson [Joh73] developed a technique to separate biological motion from other visual cues, such as the silhouette or shape of a person by attaching light points to the body and recording their movements. He set the camera threshold so that everything but those light points would be uniformly black. He showed that people are able to identify human walks based on only twelve point-lights. In his first experiment, he showed about 10 steps of walking motion as point-lights to 10 naïve students. The forward motion had been subtracted, so that the pattern seemed to walk on the spot. All participants reported seeing a walking person without hesitation. The same result was achieved when reducing the length of the stimulus to 1.0 second (a little less than half a step cycle) and in a later experiment to 200 milliseconds [Joh76]. Only when the stimulus was shortened to 0.1 seconds, was recognition performance reduced to just 40% of the participants. With these experiments, Johansson showed that human motion can be recognized accurately with only a few motion cues in a very short time. His

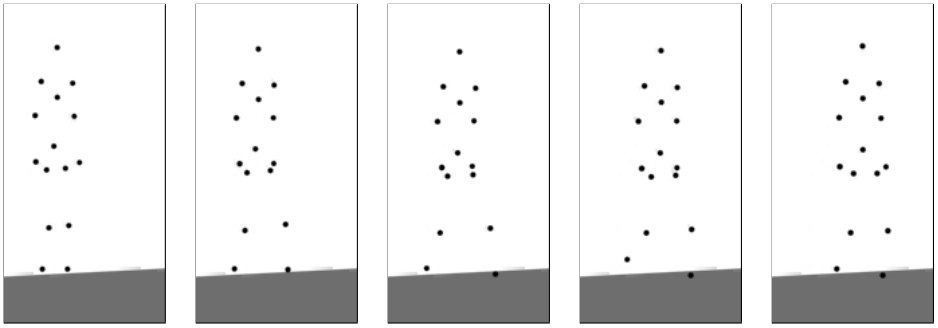


Figure 2.5: Frames from a point-light walker animation. The time span between each pair of frames is 0.1 seconds.

technique of *point-lights* to create stimuli has since been reused many times to understand how biological motion is perceived. An example of a frame from a point-light walker animation is shown in Figure 2.5.

As subsequently established in several studies, very subtle details can be distinguished and interpreted differently. Kozlowski and Cutting show that observers can recognize the sex of a character [KC77, KC78] or even the identity of a known walker [CK77] from a set of point-lights attached to their body.

Point-light displays have also been used to study the perception of emotions. Atkinson et al. [ADGY04] show that five basic emotions – anger, disgust, fear, happiness and sadness – are recognized above chance when played by a human actor. They recorded an actor playing the five emotions as a complete representation and as a view of 13 strips of white reflective tape on video. Out of those videos, single pictures were chosen. With a perceptual study using forced choice emotion recognition of the five emotions in four conditions – pictures of the acting human (full-light stills), movies of the acting human (full-light movies), pictures of the point-light representation and movies of the point-light representation – they found that full-light movies are recognized best, followed by point-light movies and full-light stills, then point-light stills. In a second study, they evaluate the perceived intensity of the emotions and draw the conclusion that the rated intensity of the motions does not depend on the shape of the character, as the point-light representation and the real human are equally rated. Further studies have analysed the recognition of emotions from body postures or body motions. Even though each study uses a different type of stimulus (Wallbott [Wal98], for example, uses video takes of actors while Coulson's stimuli [Cou04] consist of computer generated postures of a simple figure similar to a wooden mannequin), the results of these experiments are highly consistent: they con-

firm that body movements and static postures are sufficient to convey at least the basic emotions. This result motivates our study on the effect of geometry on perceived emotion (see Section 3.2.4).

The perception of hands and gestures has been studied as well, whereas in most cases hand and finger motions are treated as one. In his book on gesture and language, McNeill [McN92] thoroughly describes hand postures. However the main focus of gesture creation is on the motion of the arms, while the hands are not treated individually.

All the studies presented in this section demonstrate people's impressive ability to interpret very subtle cues in the motion of real humans. In the next section, we discuss research on the perception of virtual characters.

2.2.2 Perception of Virtual Characters

The perception of motion and motion manipulations on virtual characters has been evaluated by various researchers. In these studies, multiple short views of a single character walking, jumping or dancing, are shown. User responses are typically used to derive perceptual thresholds or sensitivity measures. Investigating human perception of correct physical laws, Reitsma and Pollard [RP03] show that participants in their study are sensitive to velocity and acceleration errors created from editing motion capture sequences containing jumping humans. Harrison et al. [HRvdP04] use a simple link structure to examine how the limbs of more realistic models may be stretched without the viewer perceiving the distortion. McDonnell et al. [MEDO09] analyse the perception of desynchronized body motions in conversations. Wang and Bodenheimer [WB04] run a user study to determine the optimal blend length for transitions between two motion clips. By comparing pairs of animations with blended transitions of different length and asking participants to determine if those animations are identical or different, they found that the just noticeable difference for motion blends is between 7 and 8 frames. They furthermore compared several techniques to compute the best blend length by asking participants which animations they judged to be more natural.

Furthermore, perceptual studies of virtual characters can help to use computer resources in an the most efficient way. McDonnell et al. [MNO07] investigate how parameters, such as the character model, motion type, cloth simulation method, linear velocity, or background complexity influence the perceived smoothness of animated virtual characters depending

on the pose update rate. In a series of psychophysical experiments, they determine that pose update rates between 16pps and 40pps – depending on the values of the parameters – are needed to guarantee smoothness. Based on these results, they suggest which update rate to use in which situation. In a later study, McDonnell et al. [MLH*09] developed a selective variation method that effectively uses the available resources to generate heterogeneous crowds. By analysing the participant's gaze, they found that people tend to first look at the head and upper torso of virtual characters in crowds.

Despite the important role of hand and finger motions in daily interactions [McN92, Nap80], we are not aware of previous studies that focus on the perception of finger motion manipulations of virtual characters.

The above cited studies analyse different types of human motions and different errors or changes in those motions, but they all point out that even subtleties in the motion of virtual characters can be recognized. Evidence from the field of neuroscience shows that real and virtual stimuli are not perceived in the same way. Different areas of the human brain are activated when presented with real and virtual stimuli. Pelphrey et al. [PMM*03] focus on the activation of neural patterns due to human motion compared to synthesized motion. They compare the reaction of viewers when seeing the motions of a virtual character, a robot, a mechanical assembly, and a grandfather clock. They find that the Superior Temporal Sulcus (STS) is sensitive to biological motion itself. Perani et al. [PFB*01] demonstrate a difference in the reaction to real and virtual stimuli by showing sequences of a real hand and two virtual reconstructions of different quality to participants. They also detect limited differences due to the degree of realism of the reproduction. Han et al. [HJH*04] analyse the differences in brain activity when participants watch cartoons or movie clips. They find that the human brain functions in a different way when interacting with real people in everyday life than with artificial characters or static social stimuli.

Not only the analysis of brain activity indicates that people process human motions in a different way than non-human motions. Kilner et al. [KPB03] find that the accuracy of participants' ability to perform a movement of their own arm differed depending on whether they were observing the motion of a human or that of a robot. According to the so-called "mirror system" theory, certain brain regions involved in executing actions are activated by the observation of actions. This would facilitate the execution of an action while observing a similar action and interfere while observing a contradicting action. Participants were instructed to move their arm either up/down or right/left, while simul-

taneously observing either a robot or a real person who performed the same or the opposite motion. The variance of the movements was analysed. The observation of another human while executing the opposite motion had a disturbing effect, while the observation of a robot did not. This demonstrates that human and robot motions are processed differently. Chaminade et al. [CFOC05] used this finding for a similar experiment. In their study, participants observed a humanoid robot moving either with an artificial motion following a sine function or mimicking a human, motion captured motion. They found that the difference between the variances occurring while performing the same and the different motion was significantly increased for the biological motion compared to the artificial motion.

These results raise the question whether the realism of a model on which a motion is displayed might influence the perception and/or interpretation of a motion. In the next section, we present studies investigating the influence of the geometric model used to display a character on the perception of motions.

2.2.3 Influence of the Geometric Model

Hodgins et al. [HOT97, HOT98] were amongst the first in computer graphics to systematically investigate the influence of the geometric model on the perception of motion characteristics. In their work [HOT98], they describe an experiment in which stimuli are generated by adding variations to a running motion. It should be noted that the original motion was simulated, not motion captured. The motion is rendered on a stick figure model and on a male polygonal model with different levels of variations added. The participants watched pairs of motion and had to rate whether these were the same or different. For three types of motion variation, subjects were better able to observe subtle changes with the polygonal model.

Reitsma et al. [RAP08] investigate whether errors in ballistic motions are perceived differently on a virtual human character than on a simple geometric form, such as a cannonball. They find differences in perception only for some type of errors, i.e. for gravitational errors but not for horizontal or vertical acceleration errors. A possible explanation for this result is that, because the change in gravity corresponds to time scaling of jump, the motions of the limbs and head provide additional information to the participants.

John and Tassinary [JT05] study the effects of shape and motion on the perception of sex, gender and attractiveness. They use silhouettes of human body shapes with varying

waist to hip ratios, from exaggerated hour-glass to tubular figures and synthetic motion restricted to two parameters (swagger for men and hip sway for women). They find that the shape of the character is more informative of the sex of the individual than its motion. This is contrary to the results of [MM94]. By applying exaggerated synthetic walks to different point-light walkers in congruent and incongruent male/female/androgynous pairings, they show that motion is more important than shape information for sex recognition.

Our investigations (see Section 3.2.4) point to an influence of both, shape and motion. Consistent with these findings, a recent study [JT07] shows that both the shape and the motion influence participants' judgement of attractiveness. In another experiment, Chaminade et al. [CHK07] investigate how the appearance of computer animated characters influence the perception of their actions. Their results suggest that the perceived biological nature of a motion decreases with characters' anthropomorphism.

The influence of the geometric model on the perception of motions is still not fully understood. One goal of our studies is to provide further insights into this research area. We perform two studies that investigate the influence of geometric models on the perceived sex and on the perceived emotion of a character (see Section 3.2.4). Furthermore, our studies in Sections 4.2 and 4.3 use different geometries. A related topic bearing many unknowns is the area of the *uncanny valley* that we discuss in the next section.

2.2.4 Hypothesis of an uncanny valley

The importance of the motions of characters is demonstrated with experiments from the field of robotics. A particularly good example is the work by Hiroshi Ishiguro who attempts to build very human-like robots. Some of his robots are that close to reality, that photos of the robot are almost indistinguishable from photos of the human original from which the robot was built. In one of Ishiguro's studies [Ish06], a total Turing test was performed by showing an android for two seconds to each of 20 participants who were told to identify the colours of the clothes. After having seen the android, participants were asked if they became aware of the android nature of the character. There were two types of android, one static and the other moving slightly. A total of 70% recognized the android nature of the static character, but 70% did not recognize the slightly moving android as being non-human.

In order to give advice for building robots, Mori posed the following hypothesis [Mor70].

He assumed that the more human-like a robot or android becomes, the more increases our sense of its familiarity. However, for androids that are very close to human-like, he hypothesized that perceived familiarity would drop. Visualizing this hypothesis by plotting *Familiarity* (y-axis) against *Human Likeness* (x-axis) creates a function that is first increasing, but then – over a limited range – drops down. The resulting curve (see Figure 2.6) is commonly referred to as the *uncanny valley*. In Mori's own words:

> *I have noticed that, as robots appear more humanlike, our sense of their familiarity increases until we come to a valley. I call this relation the "uncanny valley".* [Mor70, p.1]

He further suggested that the waveshape of the curve might be increased through motion. This assumption would suggest that humans are more sensitive to errors in motions if the character and the motions are very human-like (see Figure 2.6). For example, a humanoid robot with some anthropomorphic features may appear familiar enough to be appealing, but different enough from a real person so as not to cause fear. On the other hand, a highly realistic android (or indeed a wax-work model) may look so lifelike that it might be convincing at first glance. However, when it moves in a not quite human-like manner, it can remind people of the living dead, and hence may fall to the bottom of the hypothesized valley.

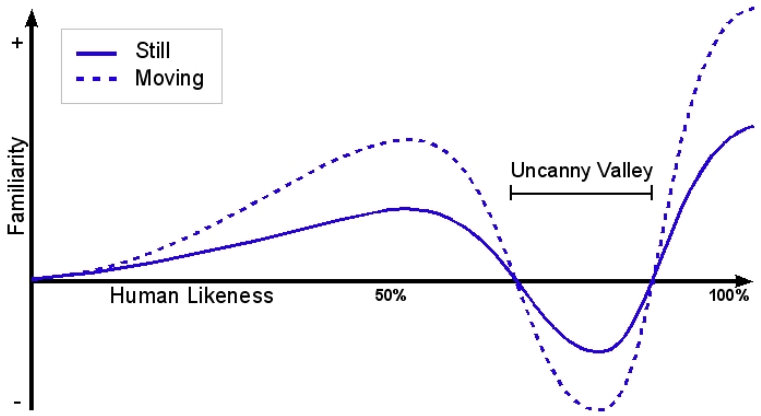


Figure 2.6: The hypothesized graph for the uncanny valley response (redrawn according to [Mor70]).

More recently, researchers attempted to prove, negate, or explain this hypothesis. MacDorman and Ishiguro [MI06] used a series of eleven pictures as an input. They showed a morph from a mechanical robot on the first picture to an android on the middle picture to the android's human model on the last picture. Participants had to rate the pictures on

the three scales strange to familiar, eeriness, and mechanical to human-like. The results can be interpreted as a confirmation of the uncanny valley theory, although the question remains whether this result arises from an imperfect morph or from a true instance of the uncanny valley. For example, double images of the eyes sometimes arise in a morph between two images that are not sufficiently well aligned. Hanson and colleagues [HOP*05] performed a carefully designed morph between a cartoon rendering of a female character and a similarly attired and posed image of a woman. They found no uncanny valley in a web-based survey of the "acceptability" of the images.

Schneider et al. [SWY07] used pictures of 75 virtual characters randomly selected from a larger pool from video games and animation. Participants had to rate how attractive they perceived the character to be and how human the character looked. The answers were graphed and the relationship could be interpreted as an uncanny valley. Contrary results are obtained by Tinwell and Grimshaw [TG09a] who use a set of 15 characters. They conclude that perceived familiarity is not only dependent on appearance and behaviour and thus the uncanny valley hypothesis is oversimplified.

Because of the difficulty of developing controllable stimuli, the existence of the uncanny valley for motion has been tested in only a few experiments [Mac06]. Using a set of 14 video clips of moving robots or androids, MacDorman [Mac06] did not find an effect similar to the hypothesized graph. The videos exhibited a wide range of behaviours with some robots only walking and other characters talking. For some androids the full body was visible, whereas for others only the head was displayed. The environments varied as well. The stimuli are therefore hard to compare and the validity of the experiment is doubtful. Using videos of 17 robots ranging from wheeled soccer playing robots to androids and one human video, Ho et al. [HMP08] explored various possible axes for the uncanny valley graph, including eerie, creepy, and strange. Correlations between perceived eeriness, voice and lip synchronization, and facial expression have also been found, where the lack of synchronization was particularly disturbing [TG09b]. These studies are limited in that they most often use images of existing robots or virtual characters and as such do not span the space of designs but just evaluate the efforts of skilled robot or character designers

Researchers have also explored different measures for assessing participants' responses to images or video suspected to span the uncanny valley. MacDorman developed an elegant experiment using terror management techniques as his metric [Mac05]. In this study, he found that a photo of an android elicited a heightened preference for worldview supporters

and a diminished preference for worldview threats relative to a control group that viewed a photo of a woman. These results indicate that perhaps an eerie robot elicits an innate fear of death or disease (as measured by the terror management metrics). Patterns of gaze fixation have also been used to determine whether an android is responded to in a similar manner as a human or a non-humanoid robot. Shimada et al. [SMII06] measure the gaze of a participant answering questions from either a real human, an android, or a robot. They divided the field of view of the participants into eight directions and compared the percental duration of gaze in the segments. They found that the gaze patterns differ significantly when answering the robot or the two other questioners. The android was treated as a human leading to the conclusion that the gaze behaviour of a person could be used to investigate the human likeness of an android, which could be applied for virtual characters as well.

Although Mori's theory has still neither been proven nor rejected, there is evidence that people's sensitivity to small errors in motion is higher in the case of a very human-like character and very human-like motions. In this research we investigate different kinds of unrealistic motions that occur with virtual anthropomorphic characters. Our aim is to discover which effects such motions have on an observer. How much can natural motion be altered before an observer notices it? Are some errors in motion less uncanny than others? And finally, which are the main points an animator has to keep in mind to create believable, lifelike virtual humans?

Chapter 3

Methods

The results and contributions presented in this thesis heavily rely on the use of motion capture technology and perceptual experiments. In this chapter, we describe the techniques and methods on which our research is based. We detail how we generated our stimuli and data and how we evaluate it. To create valuable animations for perceptual experiments and realistic hand motions for analysis, we exploit up-to-date optical motion capture techniques. After giving an overview of the animations that we generated in the course of this work, we provide an introduction of hand anatomy and present the skeleton with which we approximate this anatomy. Then we detail how we created hand and body motions, describing camera configurations, marker sets, the capture process and the post-processing pipeline. Finally, we focus on facial and eye movements and their particular challenges.

In the second part of this chapter, we describe the research methods on which our perceptual experiments are based. The awareness that we need more accurate methods to evaluate the quality of results and the effects of algorithms in computer graphics has steadily increased in recent years. We do not only want to measure how quick an algorithm performs but we also want to be able to know its effect on the audience. However, people's reactions to animations are subjective and therefore challenging to measure. To evaluate viewers' responses and to understand their underlying processes is one of the goals of perception research. The field of perception in computer graphics, nevertheless, is still in its infancy. For that reason we first introduce the basic concepts of experimental research. We then describe methods of data acquisition and data analysis related to our field, focussing on those used in Chapters 4 and 5. Finally, we illustrate these methods

by presenting two perceptual studies investigating the effect of shape and motion on the perceived sex and the impact of geometry on the perceived emotion of virtual characters.

3.1 Creating Realistic Virtual Characters

The creation of virtual characters is a complex process. As mentioned above, to generate animations as realistically as possible, we used motion capture technology. This section describes in more detail how the stimuli for this thesis were created.

The following animation sequences were produced throughout this research to examine a variety of issues:

- male and female **walks** helped to evaluate the effect of shape and motion on the perceived sex of a virtual human, see Section 3.2.4,
- with portrayed **emotions** we analysed the effect of geometry on the perceived emotion, see Section 3.2.4.
- with the vignettes *Argument*, *Bus Wait*, and *Computer Crash* (**ABC**), we investigated the effect of synchronization errors in character animation, see Section 4.1,
- the vignettes *Milk*, *Money*, and *Moving Out* (**MMM**) were used to study the impact of degradations in character animation, see Section 4.2,
- the **gestures** *Count*, *Drink*, *Snap*, and *Point* were created to find just noticeable synchronization errors of finger motions, see Section 5.1.1,
- finally, with clips from a **narration**, we compared the perceived quality of different animation methods, see Section 5.1.2.

Table 3.1 shows an overview of the number of sequences, their length, and the number of performers captured for each set of stimuli. The last column details what kind of motions were captured.

The remainder of this section first gives an overview of the basics of the complex hand anatomy and relates it to a typical skeleton model for character animation. We then describe the creation of movies where body and hands are animated with an emphasis on how we refined the capture process to record finger motions. We finally detail the generation of the MMM-scenes, which were the most complicated ones as they include

stimuli	# sequences	length	performers[1]	what was captured?
walks	6	2–5s	1 (6) lay	body
emotions	6	2.5s	1 pro	body
ABC	3	17–44s	1–2 (4) pro	bodies, hands
MMM	3	27–30s	2 pro	bodies, hands, faces, one eye, voices
gestures	4	4–17s	1 lay	body, hands
narration	4	3–6s	1 pro	body, hands

[1] pro depicts a professional actor or a drama student, lay means a layperson. The number of performers is specified for one sequence. In brackets, the total number of performers for all the sequences in the stimuli set is detailed if different individuals were captured for each sequence.

Table 3.1: Overview of motion captured stimuli.

the recording of one actor's eye movements and the capturing of bodies, hands, faces, and voices of two actors.

3.1.1 Hand Anatomy and Skeleton Model

Palastanga et al. [PFS06] describe "the development of the hand as a sensitive instrument of precision, power and delicacy" being "the acme of human evolution". Through a system of highly mobile joints and levers the hand can be held steadily at any point in space to operate tools efficiently. It has a rich nerve supply and is significantly represented in the motor and sensory regions of the cerebral cortex. The hand consists of the palm and five digits: the thumb, the index, the middle finger, the ring finger and the little finger or pinky. Due to its position and anatomy, the thumb can be opposed to the four other digits, which enables powerful grasping. Although, strictly speaking, the term finger is used to denote four digits and excludes the thumb, for simplicity we use this term to refer to all five digits.

The hand skeleton is attached to the forearm through the eight carpal bones of the wrist, and consists of the five metacarpals of the hand palm, and the phalanges of the digits: two in the thumb and three in the index, middle finger, ring finger and the pinky. This sums up to 27 bones in total, not counting the small sesamoid bone of the thumb.

The hand bones are connected through joints that allow them to move. For the fingers with three phalanges, the joint closest to the fingertip is called the distal interphalangeal joint (DIP), the middle one is called the proximal interphalangeal joint (PIP) and the one connecting the metacarpals is denoted the metacarpophalangeal joint (MCP) (see Figure 3.1). The thumb consists of a single interphalangeal joint (IP) and a metacarpophalangeal joint (MCP). The metacarpals are connected to the carpal bones at the carpometacarpal joint

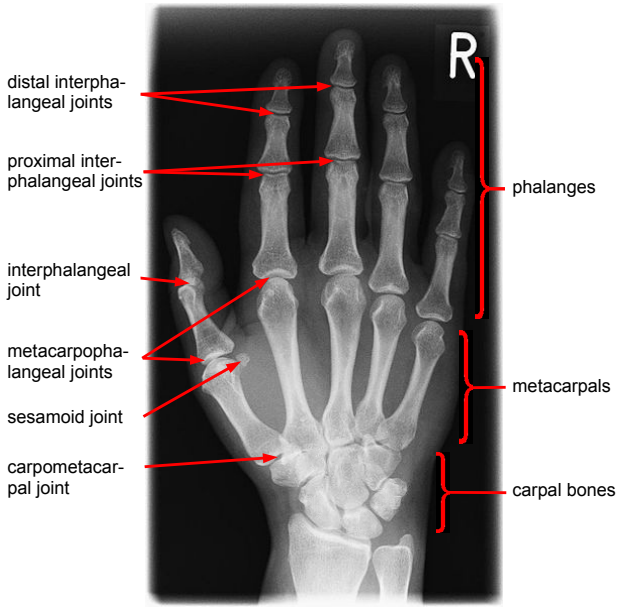


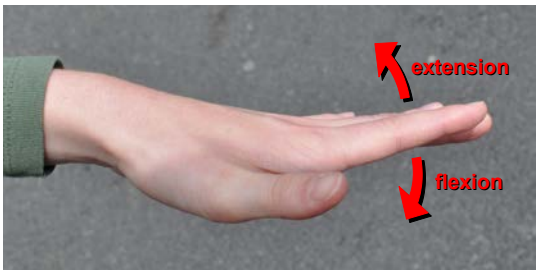
Figure 3.1: X-ray of a man's right hand illustrating different bones and joints.

(CMC).

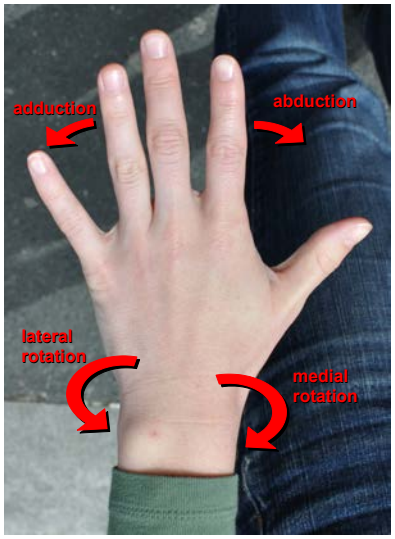
The terminology of movements of body segments is best explained with an example (see Figures 3.2 (a) and (b)). If a hand is positioned horizontally with the fingertips pointing to the fore and the palmar surface facing down (the thumb of the right hand is left and vice versa), a *flexion* of the index finger's metacarpophalangeal joint would make the fingertip point downwards. An *extension* would lift the fingertip upwards. An *abduction* moves the index towards the thumb and away from the middle finger, whereas the *adduction* moves it in the opposite way. As the index can barely be rotated around the third rotation axis in space, we describe the rotation of the whole hand. A medial rotation of the hand by 90° makes the thumb point downwards, while a lateral rotation results in the thumb pointing upwards.

Each finger can be used independently, although not to the same degree. The thumb is functionally separated and completely independent. The index finger can be controlled mostly independently, whereas the remaining three digits cannot be used on their own throughout their full range of movement, due to the arrangement of the tendons.

More detailed descriptions of the anatomy of the human hand can be found in [PFS06] or [Nap80]. To animate virtual hands, simplified models are generally used. I will first describe the skeleton model used in this research and depicted in Figure 3.3 and then compare it to other models from the literature.



side view



posterior or dorsal surface view

Figure 3.2: Woman's left hand.

Depending on how many degrees of freedom (DOF) a joint has, we model each joint as either a hinge joint (1 DOF), a Hardy-Spicer joint (2 DOFs) or a ball joint (3 DOFs), thus reducing the complex combinations of spinning, rolling and sliding motions to simple rotations around a pivot point. We use hinge joints to represent the motions of the interphalangeal joints and Hardy-Spicer joints for the flexion/extension and abduction/adduction of the metacarpophalangeal joints of the fingers with three phalanges. Their metacarpals are simplified to one structure – the palm – attached to a single ball joint, which represents the motions of the wrist. The thumb is rotated by 90° compared to the other digits, which is respected in our model. To account for the thumb's complex joint motions, the carpometacarpal joint is a ball joint whereas the metacarpophalangeal joint has only one degree of freedom.

Our model therefore has 24 degrees of freedom and takes all of the major motions of the hand that can be effected into account. We do not apply any constraints on range of motion of the joints on our model. Since we capture the movements of a real hand, no impossible motions can arise.

Parent [Par07] describes a hand model very similar to ours, the only difference being that the CMC joint of the thumb has only two degrees of freedom. This model is also adopted by El-Sawah et al. [ESGP06]. Liu [Liu08] uses a model with 27 degrees of freedom, adding one degree of freedom to the metacarpophalangeal joint of the thumb and two degrees of freedom to the metacarpals or palm. Pollard and Zordan [PZ05], instead, simply use ball joints for all the joints resulting in 57 degrees of freedom as the palm is split in three parts.

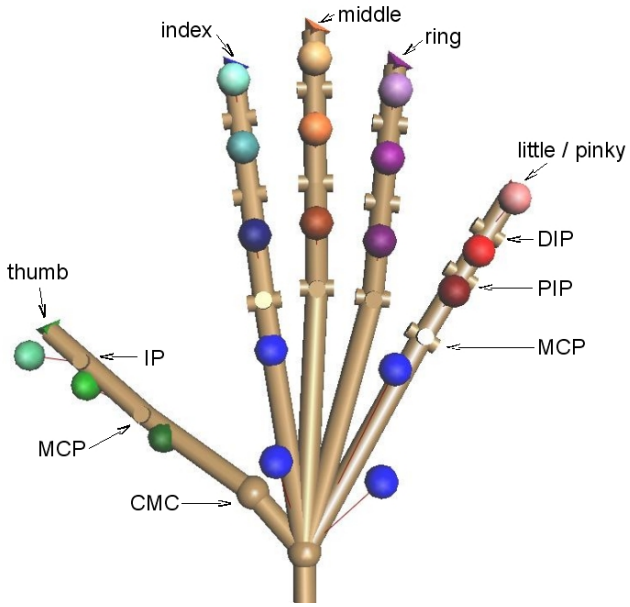


Figure 3.3: Hand model consisting of 16 joints and 19 markers.

Anatomically more correct models have been used as well. Albrecht et al. [AHS03] developed a virtual hand consisting of a modelled skeleton composed of 29 triangle meshes, the skin surface mesh with 3000 triangles, a set of virtual muscles and a mass-spring system that connects skin, skeleton, and the muscles. This level of detail is very complex and time-consuming to animate. Furthermore, if used for our models, it would make the hands stand out compared to the body. Therefore, it is not suitable for the goals of our work.

Now that we have a skeleton, we need to animate it in the most realistic way possible. This process is described in the next section.

3.1.2 Motion Capturing Bodies and Hands

In order to animate the full body and the hands of our virtual characters, we captured the motions of real performers with optical motion capture systems from Vicon [Vic11]. These systems consist of Vicon MX cameras and a data station that connects to the cameras (see Figure 3.4). Optical motion capture systems are currently the most accurate solution to capture human motion and are used in many movie and game productions.

It is challenging to capture the subtleties of the hand motions together with the movements of the body. The differences in scale and motion complexity have to be taken into account when preparing and carrying out the capture session.

For optical motion capture, retroreflective markers are attached to a performer. Cameras

(a) (b)

Figure 3.4: Hardware used for motion capturing: (a) Vicon camera; (b) Vicon datastation.

take pictures with infrared flashes, in our case with a speed of 120 frames per second, and record the resulting images. Ideally, the markers that were captured by the cameras can be seen as white discs whereas everything else is black. If the positions of the cameras are known and two cameras detect a marker, the 3D position of this marker can be computed. Once the marker positions in space are known, the markers are labelled and a skeleton approximating their motions is calculated.

The quality of a capture depends, amongst other factors, on the way the cameras are arranged – the *camera configuration* – to cover the area where the performance takes place, called the *capture volume*, and the way the markers are placed to represent the motions of the performer, referred to as the *marker set*. Careful planning of the capture session is important as cameras, capture space, and actors might only be available for a limited amount of time.

Camera configuration

Depending on the session, we had 10–18 Vicon cameras with different lenses available together with the software ViconiQ 2.0 or ViconiQ 2.5. Early tests with 6–10 cameras showed that, as a general rule, one can say, the more cameras the better.

To put these numbers into perspective: For the Hollywood production *The Polar Express* (2004), 64 to 72 mocap cameras covered a 3 by 3 metres capture area at Sony Studios (they furthermore spent $165 million, worked for three years, and hired 500 visual effects specialists). The performance capture for the movie *Beowulf* was done at Culver Studios over approximately 25 capture days starting in late 2005. The capture volume of 7.6 by 7.6 metres was covered by 250 infrared cameras. Additionally, eight camera operators recorded high-definition video footage for reference [VS07].

Although we do not have this number of cameras at our disposal, careful positioning makes a valid capture possible. This requires that one takes into account where the performers are standing, if they are moving, which direction they are facing, how tall they are, and how their hands move if the fingers are captured. We can then determine the minimal capture area needed and cover it in the best way possible with the available equipment.

Our tests of various camera configurations led to the following guidelines to capture hand motions compared to only body motions:

- a few cameras should be placed on the floor to capture the performance from below (nevertheless, cameras should never face each other),
- several cameras should be placed closer to the actors,
- the focus of each lens needs to be adapted to generate sharp enough images for the whole capture space,
- parameters, such as the gain, the strobe intensity, the circle threshold, and the threshold of each camera need to be adjusted accurately within the software.

Placing cameras on the floor is vital to capture finger motions. In many studios, the cameras are installed on a steel shell on the ceiling. With such a configuration, markers placed on the dorsal surface would not be detected by any camera during a gesture when the dorsal surface of the hand is facing the floor (thumb of the right hand points right). A closer placement increases the resolution of the cameras, which is important when the markers are small and close together.

A slightly out of focus camera might not negatively impact a full body capture, but it is crucial for a successful capture of detailed hand motions. We found it useful to glue three small markers (3–5mm), with a few millimetres of space between each of them, on a sheet of paper and to make sure that each camera could recognize the three markers and keep them apart. In the same way as an accurate focus becomes more important, camera parameters, such as those mentioned above need to be adjusted as accurately as possible. Here too we used the three small markers as a helpful device.

Examples of camera configurations following these rules can be seen in Figures 3.5 and 3.6.

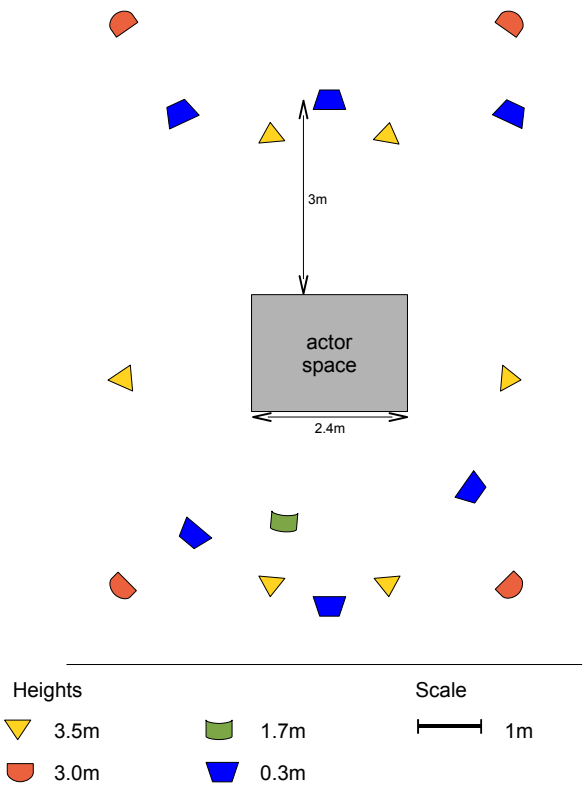
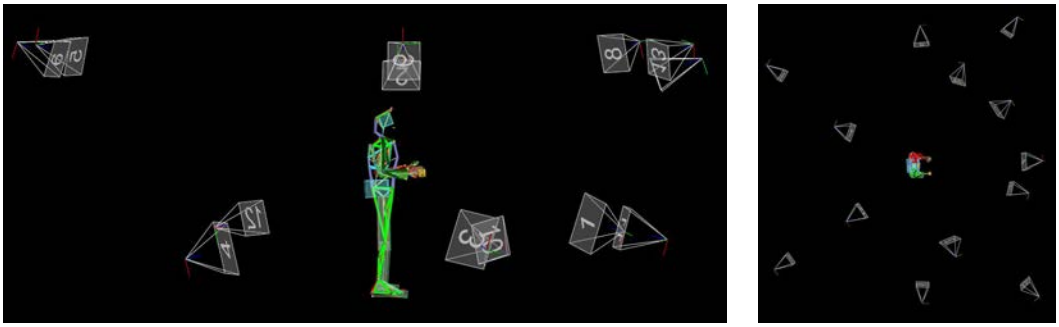


Figure 3.5: Camera position plan for the MMM capture.

Marker set

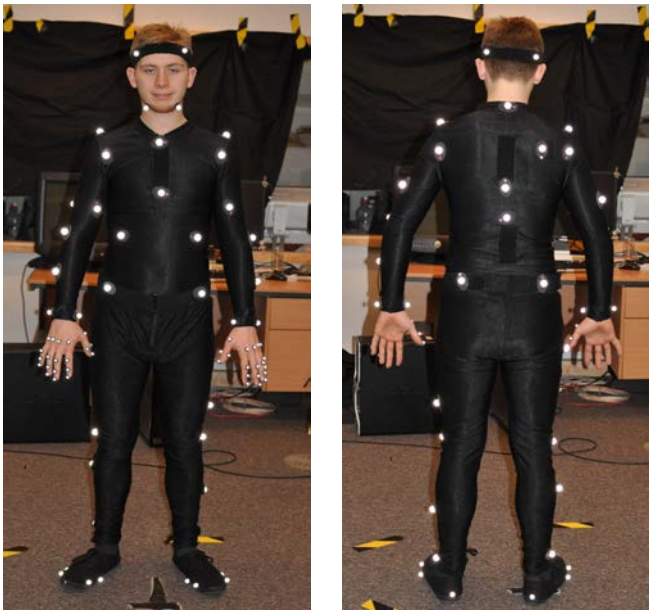
The placement of the markers determines which features of the motion are captured. Many marker sets exist for the body. An example of one that we used can be seen in Figures 3.7 (a) and (b).

To capture the hands, we experimented with marker sets consisting of between 9 and 27 markers per hand, with hemispherical and spherical markers of 3 to 6 mm diameter. Using too many markers reduces the space between markers, and two or more or them are more likely to be mistaken for one. When using too few markers, the accuracy decreases. Similarly, larger markers are visible from further apart and increase the stability of the 3D reconstruction whereas smaller markers can be differentiated even if they are very close together. A spherical shape leads to fewer gaps in the recognition, probably because the markers are further away from the skin and therefore can be detected by a larger number of cameras. On the other hand, hemispherical markers are easier to attach to the skin and are less likely to fall off during the capture process. Furthermore, we found that contrast



right side view top view

Figure 3.6: Camera configuration for the gestures capture with 13 cameras. The performer is facing right in both views.



front view back view

Figure 3.7: Marker set for the narration capture.

increases the performance of the reconstruction, meaning that black surroundings for the markers improve the results. We used tape or glue to attach the markers. When using tape, it needs to be of non-reflective type.

These recommendations are generally valid for the body as well, but their importance increases when a higher accuracy is required. In summary, our recommendations for marker sets when capturing fingers are as follows:

- use a reasonable number and size of markers depending on the task and the performance of the camera configuration,
- 22 hemispherical markers, 6mm in diameter, worked best in our setups,

(a) (b)

Figure 3.8: Marker sets of the hands used for the captures: (a) MMM; (b) narration. The differences between the marker sets are the placement and number of the markers on the thumb and the dorsal surface.

- a black surface around the markers improves the results.

The details of our marker sets were improved continuously throughout the last four years and we used the best possible configuration for each capture session based on our experience and the available markers. However, the main design stayed the same: we placed 19 markers on each hand, with the exception of the narration capture, where we used 22 markers. We attached three sensors on each finger (four on the thumb during the narration capture) and four on the palm (six for the narration session). The marker sets are shown in Figures 3.8 (a) and (b).

Capturing

For a successful capture session, the capture volume should be re-calibrated shortly beforehand, all types of reflecting objects need to be removed from the capture space, and the session needs to be carefully planned.

Depending on the task, an experienced actor might be necessary to achieve compelling results. Earlier captures of the narration with nonprofessional subjects lead to dissatisfying results. We thus hired a drama student. On the other hand, no professional actor was needed to perform a natural walk. The performers of the walks were therefore students without acting experience.

We prepared a capture plan for every session, detailing which shots are the most important, so that it was possible to spontaneously adapt to time constraints. We additionally documented the capture sessions with video footage whenever possible.

The actual capture session starts with a *range of motion*, a motion, which is used for the

calibration of the skeleton and where the joints of the body are moved exploring their full range. For the narration capture, we performed a separate range of motion for the body and for each hand, with an extended hand motion sequence to increase the accuracy of the hand skeleton calibration (see Section A).

Post-production

To post-process the movements that were obtained in the capture session, one needs to generate a skeleton, which dimensions are adapted to the dimensions of the performer. For this goal, we first build a Vicon Skeleton Template. A *Vicon Skeletal Template* or *VST* is a subject template in the software ViconiQ. It is a generic model that gets calibrated using a captured range of motion. The result of the calibration is called a Vicon Skeleton or *VSK*.

The post-production involves the following steps for each hand and the body:

1. generate VST-file,
2. circlefit, reconstruct, trajectory fit the range of motion,
3. load VST-file and label range of motion,
4. mark T-pose,
5. calibrate the VST to generate a VSK-file.

We created VST-models to define the structure of the bones and markers of each hand (see Figure 3.3). We found that generating a single VST/VSK for the hands and the body does not lead to accurate results for the hands. An explanation is that, as the errors are averaged over the whole skeleton, the errors in a single calibrated VSK are of similar size for the body and the hands. However, errors that are barely noticeable for the body can have a big impact when considered at the scale of the hands. The calibration of three separate VSTs, which are put together in a later step, solved this issue.

The construction and choice of constraints of the VST for the hands was continuously improved during this research. In step 2, the 3D positions are automatically computed by the software ViconiQ from the raw camera data, and trajectories are formed based on the representations in the single frames. This step also needs to be adapted for the purposes of motion capturing fingers by adjusting the parameters. In step 3, for each hand

and the body, the VST-file is loaded. One frame is manually labelled and the autolabel function and manual corrections are used to label the remaining take. Then the T-Poses are marked and the VSK is calculated.

We found the following techniques useful to improve the quality of the VSK:

- build a VST that already closely fits the dimensions of the desired VSK, taking into account the covariances of markers and joints,
- check kinematic fit,
- adjust covariances of markers and joints of the VSK,
- unlabel part of the range of motion and test trajectory labeler.

The marker and joint covariances represent the degree to which markers are allowed to move in relation to their segments, and joints in relation to their parent segments, respectively. The marker covariances should be small and not overlap. The sizes of the joint covariances need to correspond to the range of motion of those joints.

Now the actual motions can be post-processed and used to generate the animation sequences:

1. reconstruct, label, and smooth the needed motions with the corresponding VSKs,
2. fill any gaps,
3. perform the kinematic fit and export the required files,
4. plot the motion onto an appropriate model in an appropriate environment,
5. render the results.

These steps in the pipeline work analogously for the hands and the bodies, the only difference being that the number of gaps is usually greater for captured hands than for the body. The *kinematic fit* is the computation that generates the motion by adjusting the VSK to the marker positions. Parts of the labelling for the projects ABC, MMM, and gestures were done by Justin Macey (Carnegie Mellon University). A labelled frame can be seen in Figure 3.9.

Finally, the motion is exported. The walk and emotion stimuli were rendered in 3ds Max. The CSM-file exported out of ViconiQ was loaded onto a skeleton called *biped* in 3ds Max,

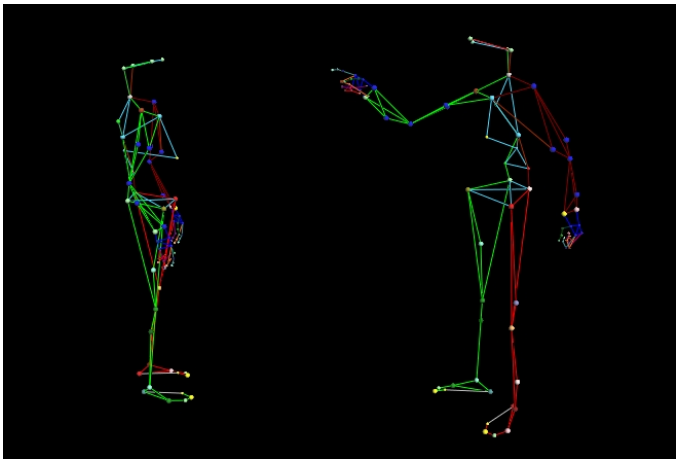


Figure 3.9: Scene *Argument* post-processed in the software ViconiQ.

a method that automatically retargets the motion to the character model. All other animations were rendered in Maya. The VSKs, containing the skeleton shape and dimensions, and V-files, having the motion information, were exported from ViconiQ and imported into Maya using scripts.

Our guidelines to capture hand and finger movements are summarized in Table 3.2. The techniques described above were used to generate the sequences for all projects listed in Table 3.1. The creation of the stimuli for the MMM-project involved additional, more complex processes, which are described in the next section.

Camera configuration
place a few cameras on the floor (avoid cameras facing each other)
place a few cameras closer to the actors
for each camera: adapt lens focus accurately
for each camera: regulate parameters (gain, strobe intensity,...) accurately
Marker set
use a reasonable number of markers, depending on the task
on the hands, 22 hemispherical markers, 6mm in diameter, worked best in our setups
a black surface around the markers improves the results
VSK
build a VST that closely fits the dimensions of the VSK (including covariances of markers and joints)
check kinematic fit
adjust covariances of markers and joints of the VSK
unlabel part of the range of motion and test trajectory labeler

Table 3.2: Summary of guidelines to motion capture hand and finger movements with an optical system and the software ViconiQ 2.5.

3.1.3 Motion Capture Including Faces and Eyes

Overview

To perform our experiments on the perception of virtual characters, we aimed to capture as detailed human motions as possible. In addition to the bodies and hands of the two actors, in the MMM set of stimuli we included the capture of facial and eye movements and sound. We produced three scenarios, each about 30 seconds long and each depicting an argument between a couple (see Section 4.2).

For this goal, we recorded:

- the face, body, and hand motions of two actors,
- video and sound,
- eye-tracker data for the female actor,
- pictures of the actors.

Figure 3.10: MMM capture session. The red lights are the real light Vicon cameras, the woman is wearing a transmitter on her back, an eye-tracker is fixed on her head.

Our equipment consisted of a camcorder, two lavalier microphones with transmitters, receivers and a mixer (see Figure 3.11), a camera, a clap, an ASL eye-tracker, and a Vicon motion capture system with six visible red and eleven near infrared Vicon cameras, which tracked the 286 retro-reflective markers attached to the bodies, fingers, and faces of the two actors (see Figures 3.10 and 3.11). Four people operated the equipment during the

session: one person operated the Vicon system and watched out for lost markers, one controlled the eye-tracker software and observed the quality of the data, one started and stopped the camcorder and monitored the sound-system, and the last one directed the actors and operated the clap that was needed to synchronize the sound and the actor's performance.

As in previous captures, we had 19 markers on each hand, and 54 markers on the body. The captured data for the hand and body markers was post-processed in the same way as described above (reconstructing the ROMs, building the VSTs, computing the VSKs, labelling the motion files, filling gaps, kinematic fit, etc.) and exported to drive two character models. The models were constructed in Maya by Moshe Mahler, 3D artist at Carnegie Mellon University, based on photographs of the real actors.

Figure 3.11: Sound system: lavalier microphones with transmitters, receivers and mixer.

Face

The capturing of the face poses similar challenges as the fingers: as the motions are more subtle, more markers on a small space are required. But there are important differences. First, the face lies essentially in one slightly bent plane, which reduces possible self-occlusions. Second, most of the facial motion is not based on bones, but on soft tissues. Thus a facial marker set consists of two types of markers: static or stable markers to measure the movements of the head and markers that measure the displacements of the skin [KW08].

The properties of faces facilitate the capturing of motion data compared to recording finger motions but make the development of a post-processing technique that takes the soft tissues into account necessary. Commercial techniques were developed for this goal in recent years. The company Mova uses reflective makeup and an optical motion capture system for their method called *Contour Reality Capture System*. It was introduced in

2006 and first used on the production *The Curious Case of Benjamin Button* (2008) by Digital Domain [Mov11]. The company Imagemetrics uses a markerless technique with video cameras where the actor needs to perform the facial animation in a limited space in which only the face is captured [Ima11].

As we needed to capture the facial animation and the body at the same time, we used a technique based on the method of Park and Hodgins [PH06]. We placed about 50 reflective small markers on each actor's face. This data was labelled and smoothed using a VSK (see Figure 3.12). Based on the results, Sang Il Park, researcher at Sejong University, Korea, created a facial geometry for each frame: the markers were segmented into seven near-rigid parts such that markers belonging to a part are roughly considered to move together (see Figure 3.13). The rigid motion of each part is approximated with six degrees-of-freedom (three for translation and three for rotation) from the movement of its member markers. The remaining local deformation of a part is then modelled with quadratic deformations and radial basis deformations and added to the rigid body motion.

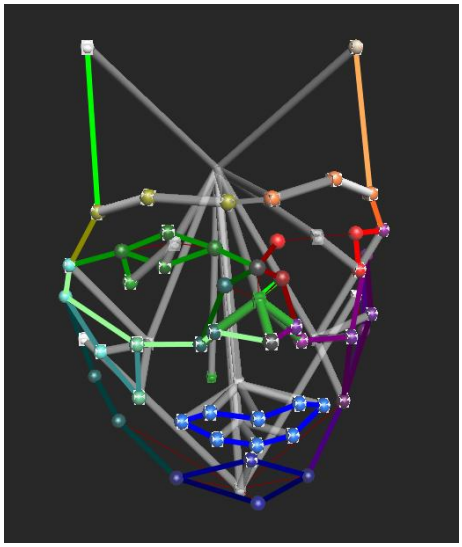


Figure 3.12: Face VSK. The bones and joints are grey, the markers are coloured and joined by sticks.

The eye contours were represented by four markers: two in the corners, one on the upper, and one on the lower eyelid. With only those markers the method of Park and Hodgins did not result in convincing geometries for the blinking. Moshe Mahler created a separate facial geometry with closed eyes and Sang Il Park blended it linearly with the resulting eye-open geometry when a blink occurs (see Figure 3.13). We then extracted the frames at which the eyelids start to close, are fully closed, start to open, and are fully open from the distance between the two markers placed on the upper eyelid and the lower eyelid,

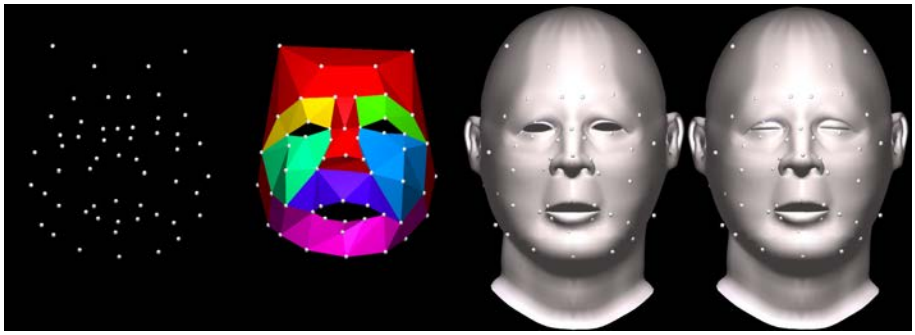


Figure 3.13: Creating the facial geometries. From left to right: facial markers, seven near-rigid parts, eye-open facial geometry, eye-closed facial geometry.

respectively, and the geometries were generated based on that data.

The hair motion was created with Maya 2008 by Moshe Mahler. After creating an initial state for the hair, the simulation was driven by gravity and the position/velocities of the head.

Eyes

The eye motion of the female actor was recorded using a head mounted ASL eye-tracker (ASL Eye-Trac 6, H6-HS Head Mounted High Speed Optics), which measures the direction of the gaze of the actor's left eye at 120fps (see Figure 3.14). To synchronize the eye-tracker data with the body motions, at the beginning of each sequence the actor gazed at a marker fixed on her hand, which she moved left and right several times. The resulting eye motion was plotted on both virtual eyeballs of the character in the final rendering.

The capturing of eye motions is not standard in the industry. In *Polar Express*, the eye movement was not captured. In *Beowulf* the electrical discharges of the muscles around the eyes were recorded to capture the eye motions [Keh07].

Our technique uses a head mounted eye-tracker. It measures the reflection of the pupil and of the cornea and calculates the direction of gaze based on this information. First tests suggested that the eye-tracker was extremely unstable when the Vicon motion capture system was in use. The detection of the corneal reflection (CR) was swapping back and forth between different spots. This happened because the infrared Vicon cameras and the eye-tracker were using similar light frequencies so that the reflections of the infrared cameras were mistaken for the corneal reflection.

Once we discovered that the ASL software scans the picture from top to bottom to find the

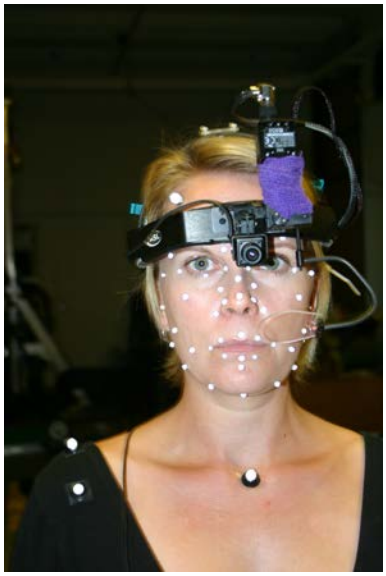


Figure 3.14: Actress with facial markers and eye-tracker.

Figure 3.15: Calibration setup. The subject sits down on the bar stool and puts his or her chin on the tripod to stabilize the head. The nine points that need to be fixated form a grid on the white board.

pupil and corneal reflection, we solved this problem with the following approach. First, we used the six near infrared cameras that did not interfere with the eye-tracker to cover the ground to be able to install all interfering cameras on the ceiling. Second, we rotated the eye-tracker camera by 180° so that the scanning would start at the bottom of the image. All remaining cameras were fixed at the ceiling at heights of 3 or 3.5 metres, with the exception of one camera on a tripod at 1.7 metres. With this technique, the correct reflection was recognized as the corneal reflection, with the exception of situations where the actor would look at spots located higher than the cameras, which did not happen in our scenarios.

To verify the performance of our approach, we recorded three sequences with an increasing

amount of eye motion in four different conditions. Each sequence was repeated twice in each condition.

The sequences were:

1. keep the head still and look at nine points on a board arranged in a 3x3 grid: small motions of the eye, no motion of the head,
2. keep the head still and look at five objects in the room, which are at the edge of the visual field: ample motions of the eye, no motion of the head,
3. move the head and look at the same five objects in the room: motions of the head and medium-sized motions of the eye.

We recorded these eye motions in four conditions:

- Vicon OFF, camera in original position: condition a,
- Vicon OFF, camera turned: condition b,
- Vicon ON, camera in original position: condition c,
- Vicon ON, camera turned: condition d.

Our assumption is that, if the eye-tracker works correctly and there is no Vicon system, for most of the successive frames the position of the corneal reflection does not vary much. Large jumps are rare. If the Vicon cameras are switched on, the reflections wrongly detected as the corneal reflection (CR) are less stable. Hence, the detected CR jumps around. So the average change of the position of the CR from one frame to the next frame is bigger. If the camera is turned around, it should have no effect if the Vicon system is switched off. When the Vicon system is switched on and the eye-tracker camera turned, we assume that the detected CR is (nearly always) the correct CR. So, the average change from one frame to the next frame should be close to the one when the Vicon is turned off.

We therefore analysed the changes of the horizontal position of the CR between successive frames (using the vertical position would have been equally possible).

We expect the following results:

1. When the Vicon system is switched off, the turning of the camera should not make any difference. That is, there will be no difference between conditions a and b.
2. When the Vicon system is on, it disturbs the CR recognition. That means the

average change between successive frames will be higher for condition c than for condition a.

3. Turning around the camera makes the CR recognition correct. That means there will be no difference between conditions b and d.

In summary, if our approach is successful, we expect conditions a, b, and d to have the same average and c to have a higher value.

The CR is not detected at all frames. The detection rate varies between 68% and 99.95% in our experimental data, depending on the type of eye motion but not on the condition. When the CR is not detected, the value of the horizontal position is given as 0. As those values would distort the average change between frames they were not taken into account for the analysis.

The results are shown in Figures 3.16 (a) to (c). The scales on the y-axis vary between the different graphs.

In Figure 3.16 (a), there is almost no difference between the trials. The task is very simple, the head does not move so the disturbing reflections of the Vicon cameras do not move either and the eye does not move much. This means that the eye-tracker performs well in all conditions. In the second type of eye motion represented in Figure 3.16 (b), the trials with the Vicon system being on and the eye-tracker camera in its original position (condition c) have higher values than the other takes, probably due to the wider range of the position of the CR. In the third type shown in Figure 3.16 (c), the difference between condition c (Vicon ON, camera in its original position) and the other takes is even larger.

Our interpretation is that, when the head moves, the reflections of the Vicon cameras change, so that the CR is more likely to be detected incorrectly. The values for condition d (Vicon ON, camera turned) are on the same level as for a and b where the Vicon system is switched off. These results are not a full proof that the camera constantly detects the correct CR but they do show that our approach works. Therefore, we used this method to capture the eye motion for our female character.

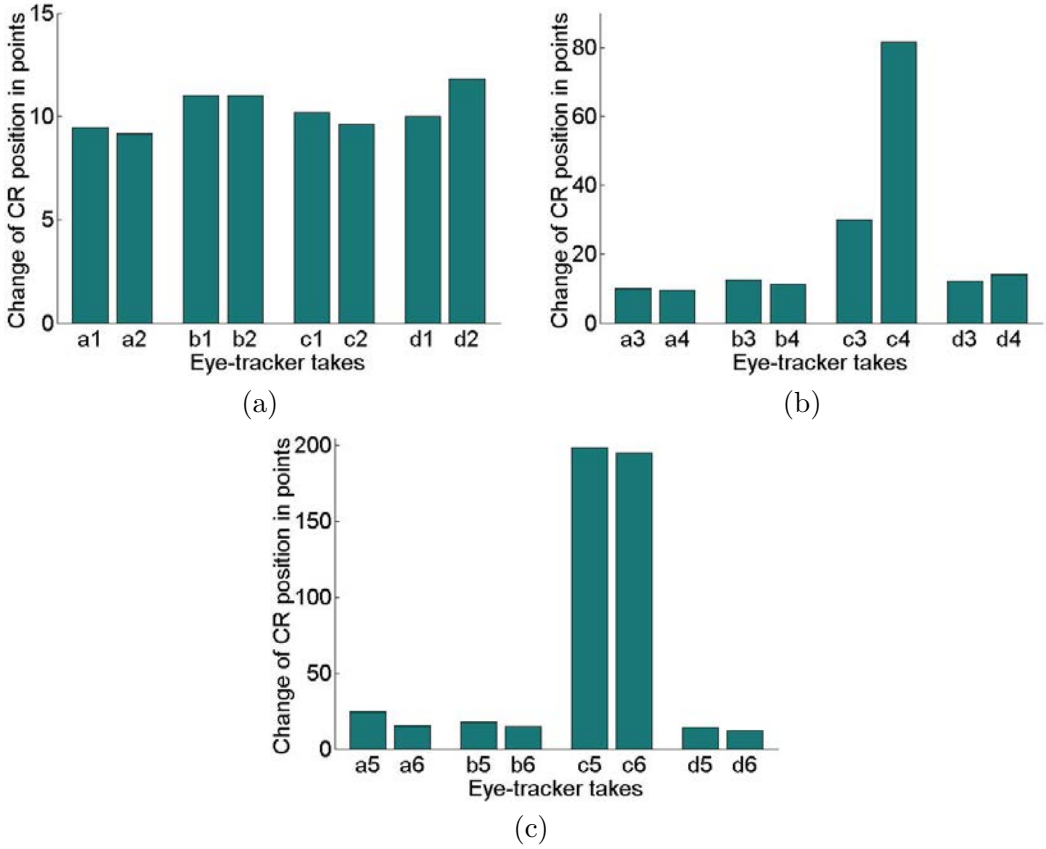


Figure 3.16: Proof of concept for capturing eye motions. Average difference of the corneal reflection's horizontal position between successive frames when (a) looking at nine calibration points on a board (in a 3x3 grid) without moving the head; (b) looking at five objects in a room at the edge of the visual field without moving the head; (c) looking at five objects in a room while head motions were allowed. There were two takes for each condition. The results are measured in points; an image captured by the eye-tracker consists of 8000x8000 points. For more complex tasks, the differences are highest for condition c (takes c3–c6), when the Vicon system is turned on and the camera is in its original position. The values for condition d (Vicon on, camera turned) are similar to the values for the conditions a and b, where the Vicon system is turned off.

3.2 Research Methods for Perceptual Experiments

Many results in this thesis were obtained by performing perceptual experiments. Having described the technical methods through which we generated our animations, we now give an overview of the research methods upon which our perceptual experiments and results are based. First, we briefly introduce the basic terminology and approach of experimental research as can be found for example in [BHFSS06] or more briefly in [Gd99]. We define terms such as *within-subjects design* or *independent variables* used in later chapters of this thesis. We then review the most established design methods to gather data, with an emphasis on methods that have been or might be useful to get further insight into

the perception of virtual humans. We thereby introduce the foundations to explain why we chose specific designs for our experiments. Examples of related work, where these techniques have been applied, are given. Finally, different approaches to statistical analysis are presented.

3.2.1 Basics of Experimental Research

The fundamental approach of experimental research states that first an experimental *hypothesis* is formulated, which in psychological research usually makes a statement about the effects of one or more events on people's behaviour. For example, we might have a variable such as the type of mesh of a virtual character and we expect that changing this variable will have an effect on the realism of the character. This property is then methodically varied in an experiment and the participant's responses are measured. We might in our example ask people to rate the realism of the character on a scale from 1 (not realistic at all) to 5 (very realistic). The *null hypothesis* states that any results found in the experiment are due to random variability in people's performance caused by unpredictable variables. To prove the hypothesis, one needs to show that the probability that our results happened by chance is so small, that this explanation is discarded. By rejecting the null hypothesis, the decision is made to accept the hypothesis that the experimental results are significant.

Different types of variables are present in an experiment: *independent variables* that the researcher manipulates (e.g. the geometry of the character), *dependent variables*, which are expected to change depending on the value of the independent variable (e.g. perceived realism), and *extraneous variables* that might affect participants' behaviour but are not the purpose of the experiment (e.g. gender, age). Various designs have been developed to carry out experiments. The most common method is to allocate participants to conditions or treatments where they are treated in exactly the same way, with the exception of the independent variable assuming different values. The *experimental treatments* are then the conditions representing the independent variable. The *control condition* is the baseline condition when the independent variable is not activated.

In a *between-subjects design* or *unrelated design*, different groups of participants are subjected to the different conditions. In a *within-subjects design* or *related design* all participants experience all conditions. These designs have advantages and disadvantages. A

between-subjects design might be necessary if the knowledge of one condition might affect the response to another condition. For example in our experiment in Section 4.1, if we ask participants about the quality of our animation, it might affect the results if we have previously shown them an animation of lower quality. Therefore, we chose a between-subjects design. On the other hand, in an unrelated design, individual differences between participant groups might affect the results. In a within-subjects design the responses can be compared to each other more easily, which has to be taken into consideration for statistical analysis.

3.2.2 Data Acquisition

As Stephens [Ste46] notes, "measurement, in the broadest sense, is defined as the assignment of numerals to objects or events according to rules". He formulates the following four types of measurement scales:

- nominal scales, where each measurement is assigned to one category, and the categories have to fulfil mutual exclusivity and exhaustiveness,
- ordinal scales, where the categories are rank-ordered, so that one category has more or less of a property than another category,
- interval scales, where the intervals between different categories are equal, enabling us to compute meaningful distances between measurements, and
- ratio scales, where a potential absolute zero value exists and meaningful ratios can be expressed.

The possible operations with those four scales are cumulative. For nominal measures, we only know whether they are equal or not, ordinal measures add the information of "more or less", interval measures have meaningful distances on top of that and ratio scale measures add an absolute zero and the possibility of expressing ratios. The type of possible operations and therefore of statistical tests and evaluations depends on the type of measurement scale.

In our research, we mostly use nominal and ordinal scales. The question, "Are you:" with the answer options "male" and "female" is a typical nominal scale. When we ask participants after showing a scene, "How angry would you rate the characters in this animation?" to choose an answer on a scale from 1 – "not angry" to 7 – "very angry", we

strictly speaking use an ordinal scale for our measurement, as we can not guarantee that the difference in angriness between 1 and 2 is interpreted to be the same as the difference in angriness between 3 and 4. However, it is widely accepted to assume these scales as being interval scales and therefore to perform statistics on them such as calculating the mean [Ste46, LCG03, BHFSS06], and we adhere to this practice.

Numerous techniques exist to measure a property, the most important ones for our purposes can be summarized as:

- physiological
- psychophysical
- questionnaire

We describe these techniques in more detail in the following sections.

Physiological data acquisition

Physiological data refers to measurements of the body. Parameters that can be measured include the muscle activity (electromyography), sweat gland activity (galvanic skin response), skin temperature, eye movements or the dilatation of the pupil, heart rate, blood pressure, respiration, or brain activity through electroencephalography (EEG) or functional magnetic resonance imaging (fMRI) [BHFSS06].

A compelling reason to use physiological data is that many types of responses are not consciously altered by the viewer. A drawback is that the interpretation is often more questionable than the accuracy of the measures would suggest. What does it imply if participants look more often at virtual character A than virtual character B? We are not able to say if this happens because the character is moving in an unnatural way or because its shape is less familiar or just because the viewer liked its clothing, based solely on this data. So without further knowledge, we cannot determine if attention would be a good variable to measure familiarity. The relationship between physiological data and the variable that we want to measure has first to be shown.

Physiological data can still give us valuable clues about participants' responses. Eye movements can be measured with a camera, an eye-tracker or by measuring the muscular activity around the eye. In our perceptual experiment in Section 4.2, we use an eye-tracker to determine how the attention of a viewer changes when watching a scene depicting virtual

characters with or without facial animation. The eye-movements of a person watching a scene typically consist of fixations during which the eyes are relatively still in one position, interrupted by saccades where the eyes perform very quick and brief motions. When analysing eye-tracker data, several properties can be considered, such as the percentage of fixations in a specific area, the fixation duration, or the location of the first fixation point. Knowing which parts of an image the attention of a viewer is drawn to can be used to investigate various questions, such as the heterogeneity of crowds [MLH*09] (see Section 2.2.2) or the human likeness of an android [SMII06] (see Section 2.2.4).

Functional Magnetic Resonance Imaging (fMRI) has also proven to be a valuable tool for gaining further insights into the perception of virtual humans. As described in Sections 2.2.2 and 2.2.3, fMRI showed that viewers have different neural activation patterns depending on whether they are watching human or non-human stimuli [PFB*01, PMM*03, HJH*04, CHK07]. The use of fMRI is unfortunately limited by the expensive equipment.

Psychophysics

"Psychophysics is the scientific study of the relation between stimulus and sensation." [Ges97]

Psychophysical methods have been developed primarily to measure properties related to sensations and perceptions. However, these experimental methods can be applied to other fields as well. Many techniques are based on Fechner's book [Fec60] and a good overview is given by Gescheider [Ges97]. One of the most common goals in psychophysics is the detection of sensory thresholds. In general, the responses to several trials are measured for different stimuli levels (intensities). One can then compute the probability of detection for each intensity and plot them to obtain a graph called a psychometric function. It is possible to present one stimulus at a trial or multiple stimuli so that the participant has to specify when or where the wanted stimulus occurred. This approach is called a forced choice technique. If two options are given, we refer to a two-alternative forced choice design (2AFC).

Other established methods to measure thresholds are the method of constant stimuli, the method of adjustment, the method of limits, and staircase methods. In the method of constant stimuli, a certain number of intensity levels for the stimuli are chosen and each is presented several times. Thus, the complete set of stimuli is determined before the

experiment, often through pilot experiments, and many trials are needed. When adaptive techniques are used, in contrast, the intensities of later trials are adjusted depending on the responses of the participant. In the method of adjustment, the participant is given direct control over the stimulus and adjusts it to the perceptible threshold. In the methods of limits, the stimulus level is first presented at a clearly noticeable level and decreased until the participant can no longer detect it. Then a level well below the threshold is chosen and increased until it is perceptible. By repeating this procedure and averaging the endpoints, the threshold is determined. The stimulus level can also be increased by a step in the next trial if it has been detected and vice versa. This is called a staircase method, as the stimulus intensities plotted over time resemble staircases. A disadvantage of adaptive techniques compared to the method of constant stimuli is that the next trial in the experiment might become predictable for the participant. Therefore multiple increasing and decreasing staircases can be randomly interleaved.

In this thesis, we used the method of constant stimuli for several experiments, see Sections 3.2.4, 4.3, 5.1.1, and 5.1.2. In Sections 5.1.2 and 4.3, we employed a two-alternative forced choice design, where participants had to choose one clip out of two in each trial. The method of constant stimuli has been applied in further studies, such as [MNO07] or [WB04] (see Section 2.2.2), to gain insight into the perception of virtual characters.

Questionnaires

Questionnaires are probably the most common data gathering method for studies involving humans, due to their versatility, simplicity and the low costs involved compared to other research tools. However, these properties do not negate the need for questionnaires to be carefully designed [KP10, SG07]. Many pitfalls exist, such as ambiguous terminology, leading questions (e.g. starting with "Would you agree that..."), or hidden assumptions (e.g. "Why do you think the man beats the woman?" assumes that the man does beat the woman). From the type of response format over the labelling of a rating scale to the order of questions, every design decision of the questionnaire has to be carefully planned, as these factors have been shown to affect respondents' behaviour.

One of the decisions that has to be made is whether to make a question open or closed. An open format allows respondents to answer in their own words whereas there is only a limited number of responses in a closed format. While a closed format can rule out

unexpected responses, an open format is more complex to analyse. For a closed question, the response format needs to be chosen, or for an open format the amount of space given.

In the questionnaires designed for the experiments presented in Sections 4.1 and 4.2, we use both, open and closed response formats, to measure different aspects of the participants' reactions. We used an open format to investigate the perceived interpretation of an animated scene by asking "Please describe what happened", providing three blank lines for participants to write down their answers. We decided against a closed format in this case as the provision of answer categories might have influenced participants' responses. Furthermore, it would have been hard to measure nuances and to take into account every kind of possible interpretation in a closed answer format. In contrast, we used a closed format to measure the emotional response to the stimuli representing the reaction to a computer crash. More precisely, we asked "How angry would you rate the character in the animation?" and provided a five-point ordered response scale ranging from 1 "not angry" to 5 "very angry". Scales of this type are commonly referred to as Likert scales and, following conventions in applied statistics, we treat them as interval scales when applying statistical tests such as the analysis of variance. In our next study, we analyse the emotional response to three scenarios depicting an argument between a couple by asking participants to rate the characters' anger levels. This time, we use an ordered 7-point response scale from 1 (not angry) to 7 (very angry) as we felt that we needed a wider range of answers.

3.2.3 Data Analysis

One of the most commonly used statistical techniques to determine whether the means of several groups differ to a statistically significant degree is the analysis of variance (ANOVA) [How08]. Its popularity can be explained by its wide range of applications: it allows to test the significance of differences between sample means, without restrictions on the number of means or the number of independent variables; it can be applied for between-subjects as well as for within-subjects designs. In addition to identifying the main effects of each independent variable, it can also be used to test for interacting effects of two or more variables. A one-way ANOVA is used when a single independent variable is tested, a two-way ANOVA involves two independent variables, and so forth. When between-subjects variables and within-subjects variables are present, a mixed-design ANOVA is used. The basic idea of the analysis of variance is to compare the amount of variance

between group means to the variance within the groups, which results in the so-called F-statistic. Based on the F-statistic, the probability of the results being random can be calculated. If this probability is less than a specified threshold, the differences observed are significant. According to the norm in psychology, we use 0.05 or 5% as a threshold for significance. We use ANOVAs in several studies presented in this thesis. For example when we analyse the effect of degradations in character animation (see Section 4.2), we show six versions of three scenes to groups of participants and analyse their ratings of the plausibility of the events in the scenario with a three-way ANOVA with the factors Condition (six samples), Scene (three types), and Gender (two types). When an ANOVA shows that there are main effects and/or interactions, a post-hoc analysis can be used to determine on which particular variable (or variables) the groups differ. We use the Newman-Keuls test, which compares each pair of means to a critical value based on the ranking of the means of the groups (more specifically how many steps the means of the pair are apart), the groups' sizes, and the significance threshold of the test.

For the analysis of free-form text answers, we use quantitative content analysis [BHFSS06], which generates numerical values from the text by elaborating a coding frame for each question. A coding frame is a set of categories into which the answers are allocated. Each answer is allocated to exactly one category. That means that the categories have to be exhaustive (every answer is assigned to a category) and exclusive (no answer is assigned to more than one category). To satisfy the exhaustiveness of the coding frame, we added a category "other" for items that would not fit into any category. Of course, one has to take care that not too many items are assigned to this category as this would otherwise destroy the purpose of the analysis. This procedure results in a frequency table, also called a contingency table, for each question. We then perform a log-linear analysis of frequency tables, which computes a statistic based on comparing the expected cell frequencies and the observed cell frequencies. Again, this statistic leads to a probability estimate and if this value is below 0.05 or 5%, we know that the effects observed are significant.

For our evaluations, we used the software Statistica 7.1 by StatSoft and spreadsheet programs as OpenOffice Calc 2.0.4 and Microsoft Excel 2003.

3.2.4 Exemplary studies

In this section we illustrate our research methods by describing two perceptual studies that we ran in cooperation with other researchers. The first one analyses the effect of shape and motion on the perceived sex of a virtual human and the second one studies the impact of geometry on the perceived emotion of a virtual character.

Effect of Shape and Motion on Perceived Sex

In this perceptual study we analysed the impact of motion and form of a virtual character on the classification of a walking motion as being male or female.[1] The results will be useful to improve the realism of real-time crowd simulation.

Four different virtual models were used: a highly realistic virtual woman, a highly realistic virtual man, an androgynous humanoid representation and a point-light walker (see Figure 3.17). The walking motions of six students – 3 men and 3 women – were captured. We used 3ds Max to plot each of the motions of the six students on each of the four characters. In addition, we synthesized three different neutral walks in 3ds Max with slight variations in the step length and applied each of the walks to each of the four characters. We showed each motion clip twice, which resulted in 72 animations of about 3.5 seconds each.

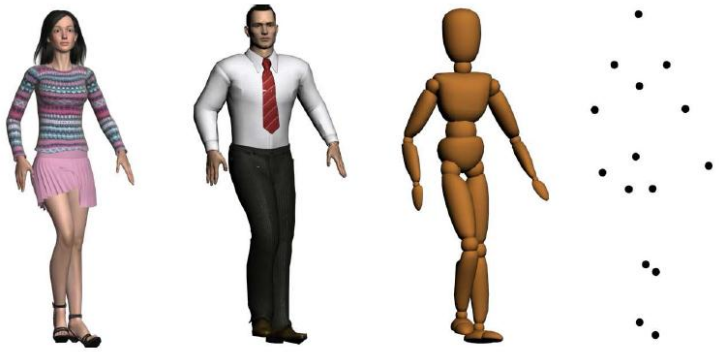


Figure 3.17: The four model representations from left to right: virtual woman, virtual man, androgynous figure (Woody), and point-light walker.

Forty one participants, divided in two groups, viewed the movies in two different randomized orders. Between each motion clip the participants evaluated the character on an answer sheet using a five-point scale: 1: very male, 2: male, 3: ambiguous, 4: female, or

[1]This study was carried out in cooperation with Dr. Rachel McDonnell and Prof. Carol O'Sullivan (both GV2, Trinity College Dublin), Prof. Jessica Hodgins (Carnegie Mellon University, Pittsburgh) and Dr. Fiona Newell (Institute of Neuroscience, Trinity College Dublin).

5: very female. Based on respondents' evaluations we showed that both, form and motion, have an impact on the perception of the sex of a character: the characteristics of a motion captured walk dominate over the model used whereas for a neutral walk the form of the model determines perceived sex.

In a second experiment we looked closer at the influence of body shape, in order to determine if additional indicators of sex in the shape will change the overall perception of the character. We used the same man and woman models as in the previous experiment wearing both jeans and a grey t-shirt, and varied their shapes (see Figure 3.18). To create a very exaggerated female body shape, we increased the size of her breasts and hips, in order to give her an hour-glass figure with a large hip-to-waist ratio (Woman 3). We applied equivalent alterations to the man model by increasing his shoulder, waist and neck width to create a more exaggerated male body shape (Man 1). We created more androgynous versions of the models, by decreasing the features we increased previously (Woman 1 and Man 3). Finally, we created intermediate woman and man versions by morphing the exaggerated and androgynous woman shapes and the exaggerated and androgynous man shapes respectively (Woman 2 and Man 2). Each of the nine different motion types from the previous experiment (3 male, 3 neutral, and 3 female) were applied to each of the body shapes (6), with two repetitions for each condition, resulting in a total of 108 clips. Three groups of participants each viewed a different random playlist on a large projected display. Thirty-eight participants (33m, 5f) took part in this experiment. As before, they categorized the character on a five-point scale, which ranged from very male to very female.

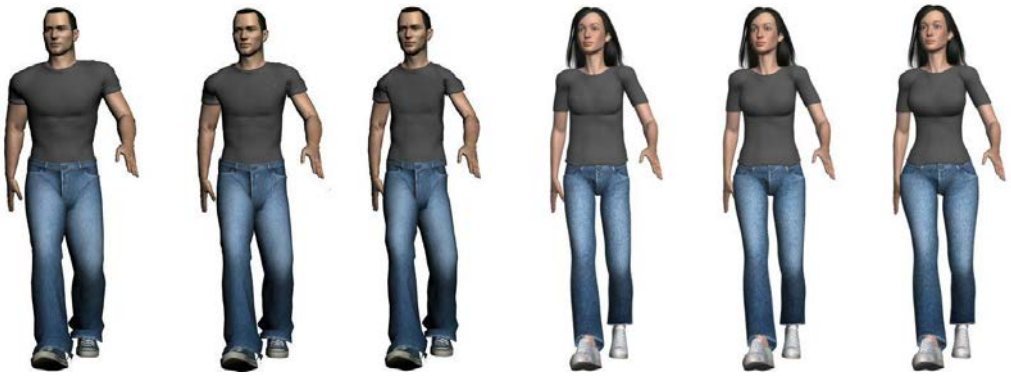


Figure 3.18: The six body shapes from left to right: Man 1, Man 2, Man 3, Woman 1, Woman 2, and Woman 3.

We found that adding stereotypical indicators of sex to the body shapes influenced sex perception. In general, the more exaggerated the male or female body shape appeared,

the more male or female respectively the character was judged to be.

Effect of Geometry on Perceived Emotion

To increase our knowledge about the influence of the appearance of a virtual character on the perception and interpretation of motions, we conducted a study focusing on the effect of a character's 3D model on the recognition of their emotions.[2]

We used the captured motions and the videos of an actor portraying six basic emotions: sadness, happiness, surprise, fear, anger and disgust. A pilot video capture with nonprofessional actors showed a lack of expressiveness in the portrayed emotions and was thus considered unsuccessful. Therefore, three actors were captured performing the six emotions up to 10 times each, out of which three versions were chosen. To choose the final motions, we asked 10 participants to carry out a basic categorization task. Each participant saw three versions of each actor performing each emotion and had to indicate which out of the six basic emotions was being expressed. We chose the actor with the best overall performance and for each acted emotion the one with the highest rate of recognition. Rates ranged from 87% to 100% identification accuracy.

In addition to the video of the real actor, his actions were applied to five virtual body shapes: a low and high resolution virtual counterpart, a cartoon-like character to see if this would be perceived to be exhibiting more positive emotions, a wooden mannequin as a completely neutral representation, and a zombie-like character which might be seen in a more negative light (see Figure 3.19). We aimed to examine the effects of body representation and motion, and did not want to confound this with expressive facial or hand movements. Hence, the faces and hands were blurred in all videos.

We used a between-groups design, so that a single participant saw the six videos with the same randomly chosen character. Each participant was then asked to judge the intensity of 41 more complex emotions, such as boredom, hope, jealousy, or shyness, that they felt the character was displaying on a 10-point scale ranging from "not at all" to "extremely". To ensure that the meaning of every emotion was understood correctly, we went through the list together with each participant and defined the more complex emotions.

We found that body representation had little or no effect on people's perception of the

[2]This study was carried out in cooperation with Dr. Rachel McDonnell and Prof. Carol O'Sullivan (both GV2, Trinity College Dublin), Ms. Joanna McHugh, and Dr. Fiona Newell (both Institute of Neuroscience, Trinity College Dublin).

Figure 3.19: The six model representations from left to right: real video, high resolution virtual male, low resolution virtual male, wooden mannequin, toon, and zombie.

emotional content of the stimuli, which leads us to conclude that a) relative to real footage, captured body motion is as effective at depicting emotional body language, b) people's perception of emotional motion is very robust and unaffected by a character's physical appearance. These results indicate that when realistic human body motion is used, it is the motion and not the body representation that dominates our perception of portrayed emotion.

Chapter 4

Perceptual Studies

In this chapter, we present a set of perceptual studies investigating the effect of errors in the animation of virtual humans. We consider very lifelike characters and create our stimuli to be as realistic as the state of the art allows. Thus, all our virtual humans are animated using motion capture technology. We present our approach of using *vignettes*, i.e., animated movies that convey emotional content. This is in contrast to previous studies in the perception of virtual humans, which have tended to use very short and simple videos with little or no affective content. The usage of vignettes as stimuli enables us to study higher level cognitive responses, such as the emotional reaction to or interpretation of a scene, which would not be possible with only the few seconds of animated content commonly used, for example, to determine perceptual thresholds. Furthermore, our approach reflects the way we encounter movies in real life situations: we tend to watch a movie once in its full length instead of viewing short snippets of it repeatedly. It is therefore important to study people's reaction to a stimuli after a single screening instead of analysing the reactions when multiple variations of a short clip are watched. This method is therefore of interest for the production of movies or similar media.

One inspiration for our approach was Mori's hypothesis of the so-called *uncanny valley* [Mor70]. This hypothesis postulates a relationship between the human likeness and the familiarity of a robot, which is assumed by some researchers to apply to virtual characters as well. The uncanny valley has not been explored in a systematic way for animated characters so far. With our experiments we aim to contribute to a better understanding of issues related to the uncanny valley.

The experiments presented here analyse the effects of synchronization errors on emotional

content, attention to details, interpretation, and perceived quality (see Section 4.1), the effect of degradations in character animation on the same variables (Section 4.2), and the effect of a variety of body and face anomalies in virtual characters (Section 4.3).

For a better understanding, we use the terms *vignette*, *scene*, or *scenario* to denominate animated movies that convey emotional content, which, in this dissertation, vary in length between 17 and 44 seconds. In contrast, we designate shorter videos with little or no affective content as *clips* or *snippets*. Their length is mostly less than 5 seconds. The terms *movie*, *stimulus*, or *video* can apply to both of these categories.

4.1 Synchronization Errors in Character Animation

Through a perceptual experiment, we investigate the effect of synchronization errors in virtual characters on the perception of short animated movies with emotional content. We find that the perception of a character's behaviour and even the whole interpretation of the vignette may be altered by errors in synchronization.

Synchronization is the correct alignment of multiple events in time. We refer to *synchronization in animation* as the correct coordination of animations controlling different parts of one movie sequence. Carefully acquired motion capture data does allow for very lifelike movements of virtual characters but for elaborate scenarios it may be necessary to combine and synchronize animations that have been captured at different times. Errors in synchronization happen in many situations. For example, when two characters play in a scene, the available capture volume might be too small to contain the actions of two performers at once. Thus, the movements of each person are recorded separately and the animations are aligned in the post-processing of the scene. If the performers act at slightly different paces, errors are hard to avoid. Capturing hand or face motion simultaneously with whole body motion is complicated due to the higher resolution required for the smaller, more detailed regions. To expand the range of possible behaviours of a character, capture libraries are often supplemented by resequencing, interpolation, and layering. These editing processes can easily introduce visually disturbing artefacts such as interpenetrations, fast twitches, or energy violations. We concentrate on three of the most common types of synchronization challenges: hands with body [MZF06], upper body with lower body [HKG06, IF04] and one character with another [SKY07].

We investigate the effects of synchronization errors on perceived emotional content, attention to details, interpretation, and perceived quality. Our hypotheses are as follows:

- H1: Synchronization errors in motion can change the perceived *emotional content* of an animated scene.
- H2: Synchronization errors in motion can reduce the *attention to details* in the background of an animated scene.
- H3: Synchronization errors in motion can change the *interpretation* of an animated scene.
- H4: Synchronization errors in motion can reduce the *perceived quality* of an animated scene.

To test these hypotheses, we generate three vignettes – *Argument*, *Bus Wait*, and *Computer Crash* – also referred to as ABC in Section 3.1, where the technical background of their generation is described. Each of these vignettes depict an event with emotional content (see Figure 4.1).

(a) (b) (c)

Figure 4.1: Frames from the vignettes: (a) Argument; (b) Bus Wait; (c) Computer Crash.

4.1.1 Method

Stimuli

Our aim was to use scenes that would immediately elicit an emotional response without too much plot development. Therefore we chose scenes close to normal life to be familiar enough to people of different ages and interests. For the *Argument* scene (17 seconds) both actors were captured simultaneously and acted out a scene in which the man is shouting and gesticulating at the woman. She first tries to remain calm, however she then loses her temper and pushes him violently away. She tries to leave but is prevented from doing so by the man grabbing her arm. In the *Bus Wait* vignette (44 seconds), the

female actor was told to behave like a very impatient woman waiting for a bus. She can be seen to check her watch frequently, sit down and get up, and swing herself forwards and backwards on her feet. Finally, for *Computer Crash* (30 seconds), the male actor was instructed that, when his computer freezes while typing something important, he should become increasingly frustrated and first try to get a response through clicking the mouse, then trying ctrl-alt-del, before finally cracking and throwing the monitor to the ground.

All characters had motion captured body movements (including the hands). For details on the capturing process, see Section 3.1. The faces remained in a neutral position without any motion. There was no audio. The recorded motions were displayed on realistic skinned models and each scenario was placed in a corresponding, high-quality rendered environment: the Argument scene in a living-room, the Bus Wait vignette at a bus station in front of a store, and the Computer Crash scenario in an office at night.

We introduced synchronization errors to these vignettes. In the Argument vignette the motions of the woman were delayed compared to the motions of the man by 0.5 seconds (15 frames). In the Bus Wait scene, the motions of the upper body, including the upper back joint and all other joints under its hierarchy (e.g. head and arms), were delayed by 0.5 seconds. This modification led to noticeable intersections of the arms and the legs when the woman was sitting down and getting up again. We removed these errors by adding a rotation offset to the arm during those motions. In the Computer Crash scene, we delayed the motions of the phalangeal joints of both hands by 0.5 seconds. We chose an error of 0.5 seconds as informal pre-tests showed that this delay is above the perceptibility threshold for all three types of error. Still, it is not as large as to be identified as synchronization error. In a follow-up experiment, we used a second modified version of the Argument vignette with only 0.17 seconds of delay, resulting in two experimental conditions (0.5s and 0.17s of error) and one control condition (no modification) for this scene. As a result we had seven vignettes: four modified and three unmodified ones.

Furthermore, to show that the chosen size of the error in synchronization is in fact detected and to find out where in the vignettes the errors are most noticed, we extracted 16 short snippets from the original and 0.5 seconds desynchronized vignettes. Four were from the Computer Crash, four from the But Wait scene, and eight from the Argument vignette.

Participants

The unmodified versions of all three vignettes were shown to 47 participants (11f, 36m), while the three modified versions were shown to 69 participants (20f, 49m). The Argument vignette that was modified by 5 frames and the short snippets of the vignettes were viewed by 64 participants (35f, 29m). All participants were students from a variety of disciplines, mostly from engineering, mathematics and science related fields, were naïve as to the purpose of the experiment and had some experience with computers. Table 4.1 shows an overview of the conditions and participants of the experiment.

group	condition	participants	date
1	unmodified CAB	16 (11m, 5f)	18.01.08
2	unmodified CAB	19 (14m, 5f)	18.01.08
3	unmodified CAB	12 (11m, 1f)	18.01.08
4	0.5s CAB	69 (49m, 20f)	18.01.08
5	0.17s A, snippets	64 (29m, 35f)	21.01.08

Table 4.1: Conditons, number of participants, and dates of each group in the synchronization error study.

Procedure

In each case, the movies were displayed on a large projected screen in a lecture theatre. We used a between-subjects design, so that no participant saw two alterations of the same vignette. The two first groups of participants each saw the three scenes – either all original or all desynchronized – in the order Computer Crash, Argument, Bus Wait.

After watching each vignette, participants wrote down their answers to a number of questions. These questions were of three types: first, those related to the emotional content, either requiring them to rate the perceived emotional state of the characters or their empathy with the characters on a 5-point scale (e.g. “How angry would you rate the character in the animation?” or “How would you rate your sympathy towards the character in the animation?”); second, those relating to facts about the scene, which required simple, precise answers (e.g. “What colour are her trousers?” or “Does the scene occur during the day or during night time?”); and third, those requiring them to describe the scene in free form text. The participants recorded their answers for each vignette directly after seeing it. They did not see the set of questions belonging to the next vignette prior to viewing it. After answering the questions for all three vignettes, participants were asked to pay attention to the quality of the animation. The three scenes were repeated and viewers

rated the quality on a scale from 1 to 5 and justified their scores. The exact wording of the questionnaire can be found in Appendix A.

The last group only viewed the Argument vignette modified by 0.17 seconds, answered the questions related to this scene and watched it a second time to rate the quality and specify the demographic data. Then, they watched the short snippets. Each clip was repeated 3 times, resulting in 48 stimuli, which were displayed in random order with a 5 seconds black screen in between, during which the participants were asked to state whether the viewed animation was modified or unmodified.

A between-subjects design for the vignettes was necessary for the purpose of our experiment. Seeing a scenario twice and answering the same set of questions could have had the following consequences: First, participants might want to stay consistent in their ratings. If they rated the woman as very angry in one condition, they might adapt their rating in the second condition. They would as well have a preconceived idea of their answer before watching the second version. Second, quality ratings of different conditions would be rated relative to each other. If participants see a vignette, they might say this vignette has a very good quality. When they see a scene with higher quality just before, this might affect their rating. Third, once the factual questions are known, participants could specifically look for the answers the second time they watch the vignette.

4.1.2 Results

We used between groups ANOVAs to analyse the five-point scale answers. To investigate H1, i.e. that errors in motion can change the perceived *emotional content* of a scene, we analysed the related questions: four for the Argument vignette, three for Bus Wait, and two for Computer Crash. The ratings were not significantly different for the two versions of the Computer Crash or the Bus Wait vignette. However, there were significant differences in the ratings of the Argument vignette (see Figure 4.2). The man was rated as significantly more angry in the version desynchronized by 0.5 seconds than in the unmodified version ($F(1,113)=5.8$, $P<0.02$), whereas the woman was rated as significantly less angry with 0.5 seconds of error than in the 0.17 seconds and original versions ($F(1,131)=11.85$, $P<0.002$ and $F(1,114)=25$, $P<0.001$, respectively). The reaction of the man was interpreted as significantly less appropriate for both the 0.17 seconds and 0.5 seconds versions than for the original ($F(1,112)=10.3$, $P<0.01$ and $F(1,105)=18.27$, $P<0.0003$,

respectively). That confirms H1 insofar as that the emotional content of the Argument scene was changed significantly by the errors in synchronization. The average ratings of the Argument vignette modified by 5 frames were mostly between the averages of the two other versions, which suggests that differences in interpretation increase with bigger errors.

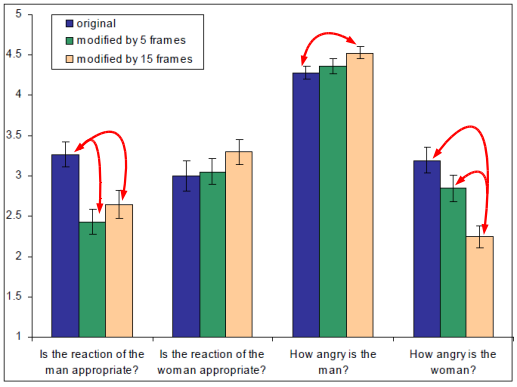


Figure 4.2: Ratings related to the perceived emotional content of the Argument vignette on a scale from 1 – not appropriate/not angry – to 5 – very appropriate/very angry. Significant differences are indicated with arrows.

The questions concerning observations about the vignettes did not differ significantly between the groups. Apparently, the errors in synchronization did not affect the ability of participants to notice details, which is why we can not support hypothesis H2.

However, the most interesting results with respect to the *interpretation* of the vignettes were found from examining the written descriptions. For Computer Crash, 21% of participants mentioned the character using ctrl-alt-del in the unmodified version compared to only 1% of the participants who saw the desynchronized version. Furthermore, 40% mentioned something about the computer freezing in the original version, but only 15% when timing errors were present – mostly, participants in this group deduced that the character's anger was due to something annoying he saw on screen or because he was frustrated when he could not do something correctly.

Closer inspection of the written descriptions also showed interesting trends for the Argument vignette, which are summarized in Table 4.2 and displayed in Figure 4.3. Most noteworthy is that no participants in the desynchronized group recognized the fact that he had grabbed her arm, with only a handful noticing that she tried to leave. In fact, seven participants in this group thought that he had tried to make her leave, e.g. "The woman has upset her husband /partner somehow. He tells her to get out of the house

and leave him alone." or "She had done something which severely upset him. He just found out and he told her to get out." The occurrences of these statements for the vignette modified by 0.17 seconds mostly lay between the values for the two other versions. These observations strongly indicate that the modifications to the finger motions in the Computer Crash scene, as well as the errors in synchronization of the characters in the Argument scenario, subtly altered the perceived interpretation of those vignettes, which supports our hypothesis H3.

Argument vignette	original	0.17s of error	0.5s of error
She pushed him	47%	41%	32%
He pushed her	4%	3%	0%
They pushed each other	2%	0%	9%
He grabbed her	23%	14%	0%
She tries to leave	23%	9%	4%
He makes her leave	0%	2%	7%
He tries to leave	2%	8%	3%

Table 4.2: Comment type occurrence rates for the Argument vignette for the answers to the question "Please describe what happened in that scene.". The rates are displayed graphically in Figure 4.3.

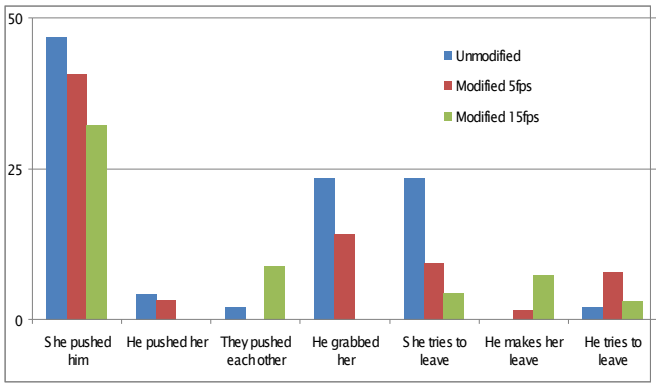


Figure 4.3: Comment type occurrence rates for the Argument vignette, graphical representation of Table 4.2. The graph shows that the occurrence rates for the version with an error of 0.17 seconds are between the unmodified version and the version modified by 0.5 seconds five times out of seven. The comment types are listed on the x-axis, while the occurrence is represented on the y-axis.

With a between groups ANOVA, we found that upper/lower body errors significantly reduced the *quality* ratings of the Bus Wait scene (see Figure 4.4). For the Argument vignette, the average quality ratings were significantly different between each of the versions, which indicates a big sensitivity even to small changes in character interaction. These results confirm H4, in that errors in the synchronization of motions can reduce the perceived quality of an animated scene. However, we found that errors in synchroniza-

tion in fingers/hands did not significantly reduce the perceived quality of the Computer Crash animation. In fact, in both versions, some participants even made comments on how realistic the finger animations were.

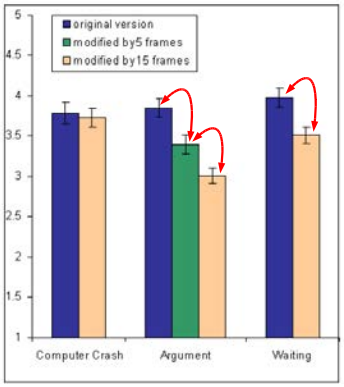


Figure 4.4: Quality ratings on a scale from 1 – low quality – to 5 – high quality – for the three vignettes Argument, Bus Wait, and Computer Crash. Significant differences are indicated with arrows.

In order to obtain greater insights into which parts of the vignettes were responsible for the differences in perception described above, we analysed the results of the *snippets* experiment. We found significant differences for the two clips showing the push and the grab in the Argument vignette, but not for the two snippets where he shouted or she reasoned. In the Computer Crash vignette, modifications were perceived for both of the snippets showing the ctrl-alt-del and the throwing. Errors in synchronization in the hip sway and the sitting down clips for the Bus Wait vignette were also perceptible. These results show that errors could be detected in all three vignettes. In the Argument vignette, they were most obvious in the part of the vignette where the contact took place.

4.1.3 Discussion

The *Argument* vignette shows that errors in synchronization can change both, the perceived meaning and the emotional content of a scene with two people. The characters' performance was quite intense and it might be an event with which participants can identify, so this vignette might have elicited stronger emotions in the participants than the two other scenes. Moreover, this vignette is the one with the most contact and the only one with interactions between two characters. This observation suggests that contacts and interactions may be a cue for emotional response and interaction. The results for the Argument vignette with an error of 0.17 seconds indicate that the changes to the

emotional content and to the interpretation increase in direct proportion to the degree of modification.

Out of all three vignettes, the fewest significant results were found for the *Bus Wait* vignette. Although the quality was perceptibly lower, no changes to the interpretation or the emotional content were perceived, which would indicate that such a vignette may not be effective for evaluation purposes. Possible explanations may be that the emotional content was lower or less negative, or because the main motion errors (after correcting for intersections) were within a single character rather than between two characters or between a character and the world, as in the other two vignettes.

The *Computer Crash* vignette also failed to elicit changes in emotional response, even though the character's performance was quite emotional. Perhaps the scene was viewed as ridiculous or humorous, as most people cannot imagine throwing a monitor onto the floor. However, what is interesting about this vignette is that it shows that errors can alter the meaning of an animation even when no quality reduction is perceived. Such errors would be impossible to detect using conventional methods, but our approach using vignettes exposed this effect. This result shows the importance of correct hand animation to make sure that the intended interpretation of a scenario is conveyed, even if viewers do not actively report that the motion is erroneous.

4.1.4 Conclusion

In summary, we found that errors in synchronization can change the way in which virtual characters in realistic scenarios are perceived. They can alter the emotional content and the meaning of an animation even without affecting its perceived quality. Therefore it is not enough for a motion editing technique to produce motion that looks realistic. The intended interpretation of the scene by viewers must also remain unaffected by the changes.

The vignettes proved to be a valuable medium for investigating the consequences of animation inaccuracies, in that interesting responses were elicited that would have been impossible to achieve using standard methods. This approach could be used to analyse the impact of other common motion artifacts in computer animation. It could also be used to investigate the importance of different aspects of an animated movie or game, such as the type of models used, the techniques for rendering skin or for cloth animation, or even the realism of sound effects. Such results would be very useful to further improve

the quality of movies and games with computer animated characters. Potential avenues for future work include developing an understanding of the complexity of our perception of animation and the factors affecting it. This research would help in the development of more reliable motion classifiers. Developing better metrics for evaluating the affective response of viewers to vignettes, and studies of the gaze of a viewer or analysis of brain scans while watching vignettes could be promising ways to gain further insights.

In the next section we present an experiment that takes advantage of this approach to investigate the importance of different aspects of an animated scene, such as the facial animation, the hair simulation, or the type of geometrical model used.

4.2 Effect of Degradations in Character Animation

In the previous section, we introduced a new method to analyse synchronization errors in animations: vignettes. In this section, we refine this approach and use it to investigate the influence of several aspects of character animation. One of our inspirations is Mori's assumption of an uncanny valley with respect to the relationship between familiarity, which is sometimes interpreted as emotional response, and human likeness [Mor70]. Many artefacts have been suggested as causes for the phenomenon, such as rigid, unfeeling faces or incorrect eye motion. Given our limited understanding of the uncanny valley, and indeed in the absence of conclusive evidence that it exists as described, we need methods to analyse the impact of different anomalies in human motion on the perception of a scenario.

Due to the complex generation process for detailed animations, in practice it may be necessary to sacrifice some aspects of the animation, e.g. captured eye gaze, or accept a lower quality for other components as the face. We took these issues into account when creating our stimuli. Knowing which aspects of an animation are most salient to the viewer and which ones might change the perceived content of a movie or game is highly useful for film-makers to increase the impact of their movies.

The results from our previous study showed that we were successfully able to create emotional content within a vignette. The most interesting results were elicited by the Argument vignette, which contains interactions and physical contact between two characters. Building from this result, we create three new vignettes, each telling a story about

an arguing couple. (The content of each vignette is described in more detail on page 75.) All three vignettes have increasing emotional intensities and are each approximately 30 seconds long. We degrade the original animations by deleting or modifying various components of the animation such as the face or the hair or by including errors such as desynchronization. We then, similar to the previous experiment, measure changes in the emotional content, attention to details, interpretation, and quality with a more detailed questionnaire. Furthermore, we analyse the gaze patterns of some participants watching the vignettes.

In contrast to previous experiments reported in the literature, in our approach we evaluate the perception of full multisensory stimuli. Instead of using short snippets of animation without context, we use longer vignettes with emotional content. To further increase the quality and realism of our stimuli, we improve our methods to generate animations as described in Section 3.1.3 and include sound, a background story, and complex animations with hair simulation and captured body, hand, face, and eye motions. We show that there are interaction effects between different factors of our experiment, and that the influence of sound is very significant in our scenes. Furthermore, we found that degrading animation quality can change the perceived quality of animations. Interestingly, an animation with a very simplified geometrical model was rated as high in quality as the Original condition, whereas other degraded animations' quality was rated as significantly lower.

Our hypotheses are identical to those in Section 4.1 with the synchronization errors in motion replaced by degradations in character animation. That means that we assume an effect of degradations in character animation on the emotional content, the attention to details, the interpretation, and the perceived quality of an animated scene.

4.2.1 Method

We collected data in two ways: First, we showed the stimuli to groups of participants in classrooms (referred to as *Group Experiment*) and asked them to fill out a questionnaire. Second, we tracked the eyes of single participants while they were watching the movies (see Figure 4.7) and asked them to complete the same questionnaire (*Eye-Tracker Experiment*).

Stimuli

We created three vignettes, each approximately 30 seconds long and depicting an arguing couple with increasing emotional intensities (see Figures 4.5 (a) and 4.6 (a) and (b)). In *Milk*, the woman gets angry about the man not leaving any milk for her breakfast. The *Money* argument is about money spent in an unwise way by the woman (at least according to the man). In *Moving Out*, the man, who does not live in the apartment any more, comes back to the house to find that the woman probably has a male visitor. We aimed to develop three scenes with increasing levels of emotional content, using emotions from the same family for all three scenes: types of anger. Between each scene, time has elapsed, which is shown by the changes in the clothing of the characters and the decor of the room. In the Milk scene, there is a heart-shaped framed picture of the couple on the fridge; in Money, the picture is half hidden by a calendar; and in Moving Out, it has completely disappeared. Furthermore, items relevant to the current situation are displayed. In Milk, cereals, a bowl, and a mug indicate that it is breakfast time; in Moving Out the rose and wine glasses allude to a romantic evening.

The virtual characters were animated using motion capture data of the bodies, faces, and hands of two professional actors. Moreover, their voices were recorded and the geometries were modelled based on the real actors. Finally, the eye motions of the female character, who was always the one facing the camera, were captured and her hair was simulated. The animations were rendered in a kitchen environment. The creation of the vignettes is described in more detail in Section 3.1.3.

We created six conditions for each scene: Original (OR), No Face (NF), Low Face (LF), Desynchronized (DS), Rigid Hair (RH), and Low Body (LB). The Original condition includes all recorded and simulated motions. In the No Face condition, the faces and eyes of both characters stayed rigid in an average, neutral shape and position during the whole scene, whereas the Low Face version kept one degree of freedom in the facial geometry – the opening and closing of the jaw – enabling basic lip motion and the eye movements of the female character. In the Desynchronized condition, we delayed the body and hand motions of the male character by 0.5 seconds, resulting in the characters being slightly out of sync. His facial animation was not moved in time so as to keep the lip synchronization with the audio track. The Rigid Hair vignette was rendered without hair simulation for the woman, so that her hair remains rigid. Lastly, in the Low Body vignette, the detailed geometries

of both characters are replaced by very simple models as can be seen in Figure 4.5 (b). This simple geometry only represents the body motions, whereas the movements of the fingers, faces, eyes, and hair were discarded. In the Eye-Tracker Experiment, we added one more condition, Audio Only (AO), where the characters are completely removed from the scene. Hence, in this condition participants only see the virtual kitchen and hear the voices of the actors. Adding a condition is easily possible in a between-subjects design. All videos had a resolution of 1280x720 pixels and a frame rate of 30fps.

(a) (b)

Figure 4.5: Frames from the vignette Milk, portraying an argument between a man and a woman over a shortage of milk: (a) Original version; (b) Low Body condition with simple geometrical models instead of detailed meshes for the characters.

(a) (b)

Figure 4.6: Frames from the vignettes: (a) Money; (b) Moving Out. They portray arguments between a couple over money spent on clothes and the messy ending of a relationship, respectively.

Participants

In our Group Experiment, 359 questionnaire sheets were collected from students in the university in the conditions Original (OR), No Face (NF), Low Face (LF), Desynchronized (DS), Rigid Hair (RH), and Low Body (LB). All of our participants were students or staff of the university and between 18 and 50 years old, with the majority being younger than 30. Table 4.3 details how many participants were in each group, specifying the condition, the class or course, the order of the animations and the date of the experiment.

A stands for the Milk scene, while *B* and *C* refer to the Money and Moving Out vignettes, respectively.

group	condition	participants	class	order	date
1	OR	13	CS 4th year	ACB	14.01.09
2	OR	14	Management science	BAC	14.01.09
3	OR	13	Italian	CBA	14.01.09
4	NF	30	CS 1st year	CAB	14.01.09
5	RH	81	Law	CAB	15.01.09
6	NF	31	Classics	BCA	15.01.09
7	DS	45	Management science	CAB	15.01.09
8	LF	93	Engineers 1st year	CAB	16.01.09
9	LB	25	Management science	CAB	16.01.09
10	LB	14	CS masters	CAB	16.01.09

Table 4.3: Conditions, number of participants, classes, vignette order, and dates of each group in the Group Experiment of the degradation effect study.

Participants for the Eye-Tracker Experiment were recruited through posters and online announcements. Unlike the other experiments presented in this dissertation, this experiment took place at Carnegie Mellon University, Pittsburgh, USA. Participants were between 18 and 60 years old, with the majority being students under 30. Their backgrounds and fields of studies were very diverse. We had the following number of participants in the different conditions: AO: 25, LB: 12, NF: 16, LF: 6, DS: 7, RH: 8, OR: 19. The order of the movies was randomized for each person.

Procedure

For the same reasons as in the previous experiment, we opted for a between-subjects design. In our Group Experiment, classes of students viewed the three vignettes (Milk, Money, and Moving Out) in one of the six conditions Original (OR), No Face (NF), Low Face (LF), Desynchronized (DS), Rigid Hair (RH), and Low Body (LB). At the start of the experiment, they filled out some general information such as gender and age, and questions related to their experience with virtual characters (for full questionnaire see Appendix C). Then they watched each vignette projected on a large display. After watching a vignette, they answered several questions about it. These questions were related to the emotional content of the scene (e.g. "How angry would you rate the characters in this animation?"), their attention to details (e.g. "What kind of picture is on the front of the fridge?"), and their interpretation of the scene (e.g. "Please describe what happened."). At the end of the experiment, after they answered those questions for each of the three scenes,

they were asked to rate the quality of the vignettes ("How would you rate the quality of the animation from a technical point of view?") and to put them in chronological order according to their understanding of the story's timeline.

Compared to our previous questionnaire investigating synchronization errors, we adjusted the questions for the three vignettes to be able to draw comparisons between them, we added a question on the physical interaction of the characters as this seemed crucial in the previous study, and, finally, we added an overall question concerning the interpretation of all three vignettes. Furthermore, we opted for a 7-point scale to examine smaller differences in participants' answers.

For the Eye-Tracker Experiment, we used the Tobii eye-tracker shown in Figure 4.7, which is integrated into the computer monitor and does not require wearing a head-mounted device. The procedure was very similar to the procedure for the Group Experiment. The main difference was that before each vignette was viewed, we had to perform a short calibration (about 15s) where participants had to watch a dot move to nine points on the screen. Furthermore, only one participant was present at any time. Subjects were given as much time as they wanted to fill out the questionnaire. The experiment usually took between 25 and 30 minutes and participants were rewarded with $10.

Figure 4.7: Participant viewing the vignettes while gaze is recorded by the Tobii eye-tracker.

4.2.2 Group Experiment Results

In the next sections we analyse the results to the questionnaires in the order of our hypotheses H1 to H4: After giving an overview of our data, we first investigate changes in the emotional content, then we test for a reduction of the attention to details followed by differences in the interpretation. Lastly, we evaluate the answers to the quality of the

scenes. We analyse the effects to support or reject our hypothesis. Then we look for further effects, which give insights as to the reasons for our results.

Our results for the Group Experiment are summarized in Table 4.4 on page 80. Only significant effects are listed. The rating differences leading to those effects are then detailed in the column "Reason".

Data Overview

We are aware that groups of our participants that were subjected to different conditions have different study fields, which was unavoidable in our design within given constraints. We therefore need to take this fact into account when discussing our results.

We had to discard some of the answers to the factual questions from groups 1-4 due to incorrect ordering of the sheets. In group 4 (30 participants, NF condition) none of the factual questions could be included in the analysis. For group 1 (13, OR) the answers to the factual questions of the Milk animation were part of the evaluation. For group 2 (14, OR) the answers to the factual questions of the Moving Out scene, and for group 3 (13, OR) those of the Money vignette, were used.

Participation in our experiment was purely voluntary and there was no reward associated with it, which we made clear at the beginning of the experiment. Furthermore, the answer times were limited. This explains why several questionnaires were not filled out properly. All answer sheets were checked for their validity. A total of 56 questionnaires was discarded for the following reasons:

- participants filled out only one page per sheet (we used double sided prints), so that half of the answers were missing: 24 questionnaires
- questionnaire was not taken seriously (joking or rude answers): 11
- numerical answers missing: 11
- invalid scoring, e.g. answering "4" at every single rating question: 4
- the factual questions were swapped (questionnaires of the conditions OR and NF were not discarded due to this reason as they were swapped for all participants in the corresponding groups): 3

Question	Effect	Probability	Reason
Angriness	Scene	F(2, 482)=121.01, p≈0	Milk < Money < Moving Out
	Character	F(1, 241)=198.08, p≈0	man < woman
	Condition*Scene	F(10, 494)=2.7788, p<0.005	Moving Out DS < Moving Out LB
	Scene*Character	F(2, 482)=241.57, p≈0	Milk, Moving Out: man<woman, Money: woman<man
Appropriate	Scene	F(2, 482)=17.255, p≈0	Milk < Money, Moving Out
	Character	F(1, 241)=122.79, p≈0	woman < man
	Condition*Character	F(5, 241)=2.2761, p<0.05	DS woman << DS man
	Scene*Character	F(2,482)=40.533, p≈0	woman: Milk<Money<Moving Out, man: Moving Out<Milk, Money
	Character*Gender	F(1,241)=21.003, p≈0)	females: man < woman, males: woman < man
Sympathy	Character	F(1, 241)=136.51, p≈0	woman < man
	Condition*Character	F(5, 241)=3.5536, p<0.005	woman: DS<LB
	Scene*Character	F(2, 482)=46.027, p≈0	woman: Milk<Money<Moving Out, man: Moving Out<Milk, Money
	Scene*Gender	F(2, 482)=6.1502, p<0.005	Milk, Money: females<males, Moving Out: males<females
	Character*Gender	F(1, 241)=32.915, p≈0	females: man < woman, males: woman < man
Plausible	Scene	F(2, 482)=22.028, p≈0	Milk < Money, Moving Out
Picture (Money)	significance	p<0.05	ordering effect
Beverage (Moving Out)	significance	p<0.001	ordering effect
What happened	Scene	F(2, 460)=14.805, p≈0	Milk<Money, Moving Out
	Condition*Scene	F(10, 460)=2.1517, p<0.05	RH: Milk<Money, Moving Out
Responsible, Money	significance	p<0.05	NF: man and woman same
Man Before, Money	significance	p<0.05	ordering effect
Quality	Condition	p<0.01	NF/LF/DS<LB, NF/DS<OR, and NF<RH

Table 4.4: Summary of results of the Group Experiment in the degradation effect study.

- participants must have filled out in the wrong order (e.g. the answers to the first animation shown refer to the third animation): 2
- information for gender is missing: 1

The number of blanks were counted for each participant. Table 4.5 summarizes the number of missing answers in the answer sheets of each condition. Based on these numbers, it was decided to include all questionnaires with less than 10 unanswered questions into the evaluation.

N/A	LB	NF	LF	DS	RH	OR	total
0	6	9	6	8	4	10	43
1-5	25	24	39	25	30	20	163
6-10	1	17	8	5	10	6	47
11-15	2	2	5	3	4	2	18
>15	0	2	0	0	0	0	2
discarded	5	7	5	4	33	2	56
total	39	61	63	45	81	40	329
total ≤ 5	31	33	45	33	34	30	206
total ≤ 10	32	50	53	38	44	36	253

Table 4.5: Number of missing answers and discarded questionnaires in each condition of the degradation effect study's Group Experiment.

Table 4.6 displays the distribution of males and females across conditions, considering only the questionnaires that had fewer than 10 unanswered questions. The proportion of female participants ranges from 21.9% to 50.0% in the different groups; we accounted for this variation in our analyses by examining gender effects.

condition	LB	NF	LF	DS	RH	OR	total
f ≤ 10	7	17	16	14	18	18	90
m ≤ 10	25	33	37	24	26	18	163
total≤ 10	32	50	53	38	44	36	253
% f	21.9	34.0	30.2	36.8	40.9	50.0	35.6

Table 4.6: Distribution of male and female participants with fewer than 10 unanswered questions in each condition of the Group Experiment in the degradation effect study.

Emotional Content

To corroborate or to reject hypothesis H1, i.e. that degradations in character animation alter the perceived *emotional content* of an animated scene, we analyse the first seven numerical questions for each scene: angriness, appropriateness, and sympathy for man and woman each, and plausibility of the scene.

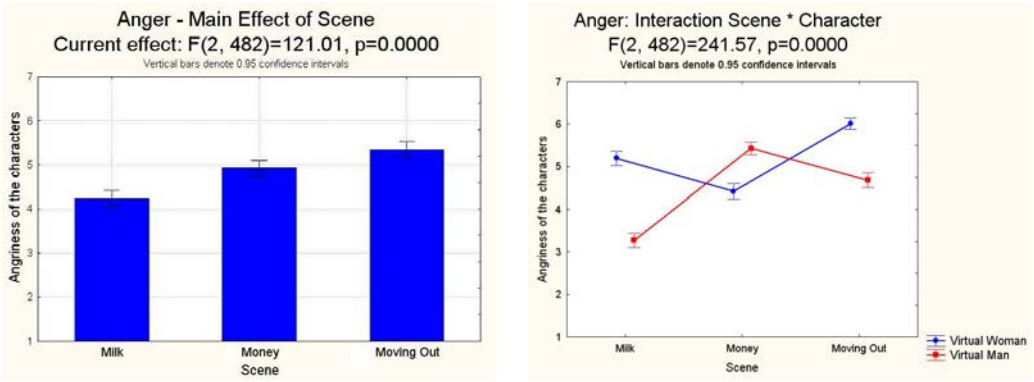
To evaluate the first three questions asked for each of the characters, we used a four-way mixed design ANOVA (repeated measures ANOVA) with the factors:

- Condition, between-subjects variable, 6 values (LB, NF, LF, DS, RH, OR),
- Scene, within-subjects variable, 3 values (Milk, Money, Moving Out),
- Character, within-subjects variable, 2 values (woman, man), and
- Gender, between-subjects variable, 2 values (female, male).

There was no main effect of Condition for angriness, appropriateness, or for sympathy. In the next paragraphs, we detail the answers to each of the questions. When interaction effects were found, we investigated further using Newman-Keuls tests for comparisons between means.

How angry would you rate the characters in this animation? There were main effects of Scene and of Character, an interaction effect between Condition and Scene, and an interaction effect between Scene and Character. The main effect of Scene ($F(2, 482)=121.01$, $p\approx0$) was due to the fact that the scenes were rated in increasing levels of anger in the order Milk < Money < Moving Out (see Figure 4.8 (a)), as intended, with all differences being significant (all $p<0.001$). On average throughout all scenes the man was rated as significantly less angry than the woman leading to the main effect of Character ($F(1, 241)=198.08$, $p\approx0$). A Newman-Keuls post-hoc analysis showed that the interaction effect between Condition and Scene ($F(10, 494)=2.7788$, $p<0.005$) was mainly due to the fact that in the Moving Out scene the LB condition was rated significantly higher than in the DS condition. In the Milk and Moving Out vignettes the woman was rated angrier, whereas in the Money scene the man was rated angrier, with all differences being significant, which results in the interaction effect between Scene and Character ($F(2, 482)=241.57$, $p\approx0$) shown in Figure 4.8 (b). In the Milk scene the woman is the one who raises her voice first (she's angry as there is no milk left), while in the Money vignette it is the man (who discovered the credit card bill). In the Moving Out scene the woman is again the first one expressing her anger; she also physically pushes the man. This explains the ratings that led to the interaction effect between Scene and Character.

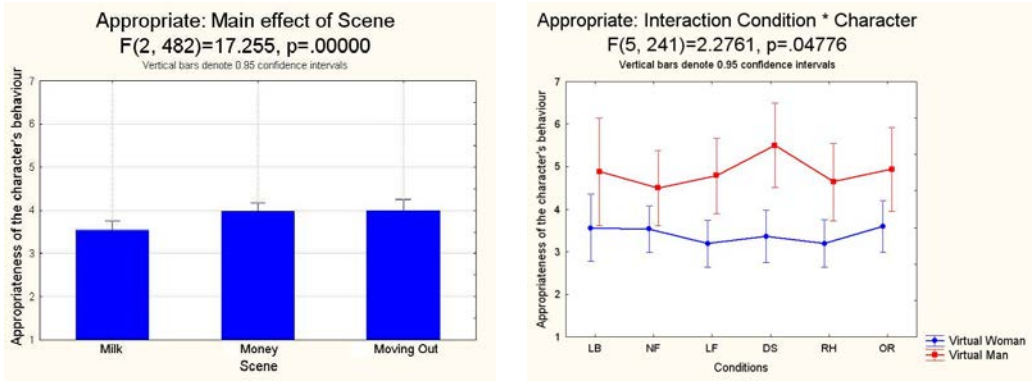
How appropriate would you rate the characters' behaviour? Our analysis showed main effects of Scene and of Character, interaction effects between Condition and Char-



(a) (b)

Figure 4.8: Results for the question "How angry would you rate the characters in this animation?": (a) Main effect of Scene: The characters of the three scenes were rated to have significantly different levels of angriness. (b) Interaction effect between Scene and Character: The virtual woman was rated as being angrier than the man in the Milk and Moving Out scene, whereas it was the other way around in the Money vignette.

acter, Scene and Character, and Character and Gender. The reason for the main effect of Scene ($F(2, 482)=17.255$, $p\approx 0$) is that the behaviours of the characters in the Milk scene are rated as significantly less appropriate than in the two other scenes (see Figure 4.9 (a)). Probably the reason for starting the argument – a shortage of milk – seemed too trivial and the woman's angry response disproportionate with the man's reaction. This was confirmed by the main effect of Character ($F(1, 241)=122.79$, $p\approx 0$), which was explained by the fact that the behaviour of the woman was rated as significantly less appropriate compared to the behaviour of the man.

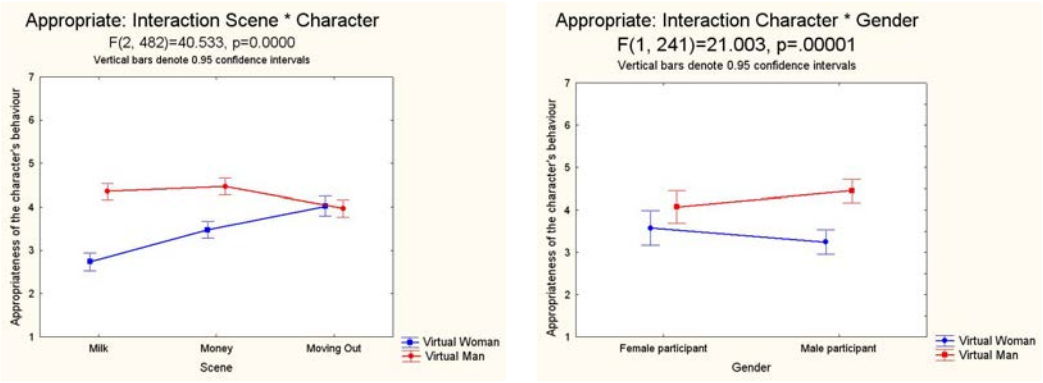


(a) (b)

Figure 4.9: Results for the question "How appropriate would you rate the characters' behavior?": (a) Main effect of Scene: the behaviour of the characters in the Milk scenario was rated as significantly less appropriate than in the two other vignettes. (b) Interaction effect between Condition and Character: The difference between the ratings of the two characters was largest for the DS conditions.

There was an interaction effect between Condition and Character ($F(5, 241)=2.2761$,

p<0.05): As Figure 4.9 (b) shows, the difference between the ratings of the two characters was largest for the DS conditions as the behaviour of the man was rated as significantly more appropriate than the behaviour of the woman. The interaction effect between Scene and Character (F(2,482)=40.533, p≈0) is explained by the fact that the behaviour of the man was rated as increasingly appropriate for the scenes Milk, Money, and Moving Out, whereas the behaviour of the woman was rated as less appropriate for the Moving Out scene than for the two other vignettes (see Figure 4.10 (a)). Interestingly, we found an interaction effect between Character and Gender (F(1,241)=21.003, p≈0), as can be seen in Figure 4.10 (b). Opinions differ as to the appropriateness of the female character. While the behaviour of the virtual woman was rated as significantly more appropriate by female participants than by male participants, the behaviour of the virtual man was rated as significantly more appropriate by male participants than by female particpants.



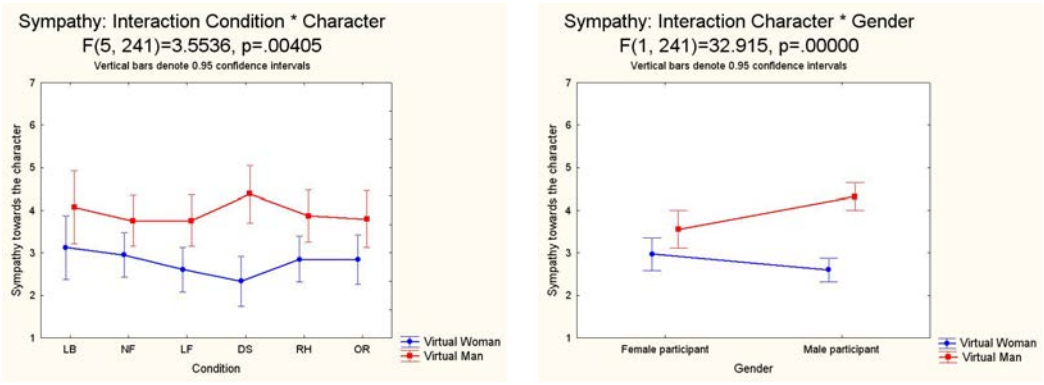
(a) (b)

Figure 4.10: Results for the question "How appropriate would you rate the characters' behavior?": (a) Interaction effect between Scene and Character: The behaviour of the man was rated as increasingly appropriate for the scenes Milk, Money, and Moving Out, whereas the behaviour of the woman was rated as less appropriate for the Moving Out scene than for the two other vignettes. (b) Interaction effect between Character and Gender: Women rated the behaviour of both character as nearly equally appropriate, whereas men rated the behaviour of the male character as clearly more appropriate than the behaviour of the virtual woman.

How would you rate your sympathy towards the characters? We found a main effect of Character, with the virtual woman rated as less sympathetic than the man, and interaction effects between Condition and Character, Scene and Character, Scene and Gender, and Character and Gender.

The interaction effect between Condition and Character (F(5, 241)=3.5536, p<0.005) can be explained by the fact that for the female character there was a significant difference

between the DS and LB condition, whereas for the male character there were no significant differences between conditions (see Figure 4.11 (a)). The interaction effect between Scene and Character ($F(2, 482)=46.027$, $p\approx0$) results from the fact that, although there were similar ratings of sympathy for both characters in the Moving Out scene, the man was rated as significantly more sympathetic than the woman in the Milk and in the Money scene.



(a) (b)

Figure 4.11: Results for the question "How would you rate your sympathy towards the characters?": (a) Interaction effect between Condition and Character: there was a significant difference between the Desynchronized (DS) and Low Body (LB) condition for the female character and none for the male character. (b) Interaction effect between Character and Gender: The virtual woman was rated as significantly more sympathetic by female than by male participants, whereas it was the other way around for the virtual man.

The interaction effect between Scene and Gender ($F(2, 482)=6.1502$, $p<0.005$) results from the fact that female participants' sympathy ratings were lower for the scenes Milk and Money than those of the male participants, whereas it was the other way around for the vignette Moving Out. Similar to the previous question, the virtual Woman was rated as significantly more sympathetic by women than by the men, whereas it was the other way around for the male character, which explains the interaction effect between Character and Gender ($F(1, 241)=32.915$, $p\approx0$, see Figure 4.11 (b)). So, here again, women tended to rate the female character more positively than men did (but not more positively than they rated the male character), while men were more in favour of the virtual man than female participants were. We can deduce from these findings that participants did respond emotionally to the virtual characters, which validates the purpose of our vignettes.

How plausible would you rate the events shown? To analyse the answers to the plausibility question, we used a three-way repeated measures ANOVA , with the factors:

- Condition, between-subjects variable, 6 values (LB, NF, LF, DS, RH, OR),
- Scene, within-subjects variable, 3 values (Milk, Money, Moving Out),
- Gender, between-subjects variable, 2 values (female, male).

No main effect of Condition was detected for plausibility either. The only effect found was a main effect of Scene ($p \approx 0$), with the Milk scene rated as significantly less plausible than the two other scenes, probably again because a shortage of milk seemed too trivial a the subject for such an angry argument.

In summary, we did not find an effect of Condition for any of the questions related to the emotional response of the viewer. However, we found interaction effects between Condition and Character for the questions concerning the appropriateness and sympathy and between Condition and Scene for the question concerning the angriness. These interaction effects were due to distinct ratings for the Desynchronized condition (DS). Nevertheless, none of our other degradations had an effect and such a result might come from the fact that we had participant groups with different backgrounds. As a result, *we can not support hypothesis H1*, i.e. that degradations in character animation alter the perceived emotional content of an animated scene.

Attention to Details

There were two factual questions for each scene to find out if participants were able to describe details of the scene that were not directly relevant to its content, and to thus support or reject hypothesis H2, i.e. that degradations in character animation change the *attention to details* in the background of an animated scene. For each vignette, our first question required a one word answer such as "wine" for the question "What kind of beverage is on the counter?" and the second one asked to choose one out of two options, e.g. "yes" or "no" for the question "Was there a vase on the shelf?". We coded the one word answers as being correct or incorrect, resulting in binary answers for all of the questions. To test for an effect of Condition, we ran a log-linear analysis of frequency tables for each of the questions with Condition and the answer as the variables to be analysed. We discarded any non-answered questions.

In summary, two of the results showed significant differences between the conditions. But these effects can be fully explained by the order in which the movies were watched by

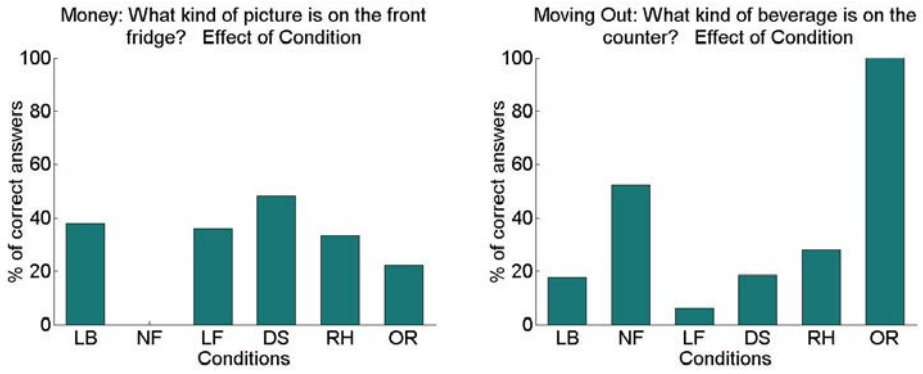
participants. Therefore, *we cannot draw any useful conclusions to support hypothesis H2.* We detail the results and coding criteria for the two questions that showed significant differences below.

Money: What kind of picture is on the front fridge? On the front fridge, there was a heart-shaped picture of the couple half hidden by a calendar depicting two dogs. The answers were coded correct if they mentioned a calendar, a heart, or dogs. Examples of correct answers are "Calendar over the heart.", "dog", "a calendar of dogs and a heart with their picture in it.", or even "rectangle of animal?". All other answers were coded as incorrect, e.g. "holiday postcard", "a nice one", or "don't know". The log-linear analysis of frequency tables indicated that a significant majority of people gave an incorrect answer ($p<0.001$). Furthermore, we found a significant effect of Condition ($p<0.05$).

To explain the effect of Condition, we examined the frequency table and the percentages of correct answers for each condition (see Figure 4.12 (a)). They show that fewer participants in the Original condition (OR) than in the other conditions knew the correct answer and none at all in the No Face condition (NF). Taking into account that the scene Money was seen first for the NF condition, as the second animation in the OR condition, and last in all other conditions, it is very likely that the effect of Condition for this question is due to ordering effects. To further test this conclusion, we performed a log-linear analysis of frequency tables without the NF condition. There is no significant effect of Condition without the NF scene.

Moving Out: What kind of beverage is on the counter? There was wine on the counter. The answers were coded as correct for every answer that included "wine", such as "red wine", "2 wine glasses" or "a bottle of wine", and incorrect for every other answer, such as "coffee", "beer" or "Don't know". The log-linear analysis of frequency tables showed that significantly fewer people answered correctly than incorrectly ($p<0.001$). Furthermore, there was a significant effect of Condition ($p\approx0$).

Here again, we look into the percentages of correct answers for each condition (see Figure 4.12 (b)). The percentage of correct answers is 100% for the OR condition, more than 50% for the NF condition and less than 30% for every other condition. The Moving Out scene was shown third for OR, second for NF and first for all other conditions. So, our conclusion is the same as for the picture question, i.e. the effect of condition can be



(a) (b)

Figure 4.12: Percentage of correct answers for (a) "What kind of picture is on the front fridge?" (Money scene) and (b) "What kind of beverage is on the counter?" (Moving Out vignette). The Money scene was seen first for the No Face (NF) condition, second for the Original (OR) condition and last for all other conditions, whereas the Moving Out vignette was seen first for the conditions Desynchronized (DS), Low Body (LB), Low Face (LF), and Rigid Hair (RH), second for the condition NF, and third for the condition OR. This is reflected in the results, which can thus be interpreted as ordering effects.

explained as ordering effects.

Interpretation

We asked for the participants' interpretation of each vignette, related to hypothesis H3, using a total of seven questions: We asked them to describe what happened in the scene in a few lines of free-form text; we wanted to know whether the man or the woman was mainly responsible for the argument; we asked what the characters were arguing about; we requested a description of the physical interaction in the scene; we generated two different questions for each vignette to inquire for example what the characters were doing before or after the scene played; and finally, we asked participants to put the three movies in the right chronological order and to tell the full story around the events depicted.

In summary, we could find trends in the evaluation of the free-form texts, but *no effective results could be found to support hypothesis H3*, i.e. that degradations in character animation can alter the perceived interpretation of an animated scene. The results to each question are detailed in the next paragraphs.

Please describe what happened Accurate rating and coding of free-form texts is challenging and can lead to exhaustive discussions. We analysed the descriptions of each vignette for the conditions Original (OR), No Face (NF), and Low Body (LB) as those were

the conditions where we expected the clearest differences. We employed two techniques: First, we asked volunteers to read the three sets of answers (conditions OR, NF, and LB) related to each vignette and to rank the three groups according to the strength of their emotional responses. Second, we asked independent investigators to rate each single answer and we analysed the results. Our goal was to have two independent evaluations of our results, which would reinforce each other. In summary, our analysis found trends of ratings, but did not provide any conclusive results.

In the first setting, seven volunteers were asked to rank three groups of answers in decreasing order of emotional response for each vignette. To allow a broad overview of the responses but to avoid independent investigators having to read hundreds of answers, which might confuse rather than give additional insight, we limited this evaluation to 25 randomly chosen answers out of each of these conditions.

The answers of the investigators show clear trends for the Milk and Moving Out vignettes. The LB condition was ranked highest by 5 out of 7 investigators for the Milk scene, while the two remaining people ranked this condition second highest. In the same way, for the Moving Out scenario, the NF condition was ranked highest by 6 out of 7 investigators and second highest by the remaining person. The second result was surprising as we did not expect the NF condition to have the highest emotional content.

In our second analysis of the scene descriptions, the level of emotion of each text is rated individually by an independent investigator on a scale from 1, meaning "not emotional at all", to 7, meaning "very emotional". Three volunteers were each given the complete set of descriptions for one vignette in randomized order through all conditions, being naïve about the different conditions. We avoided asking them to rate the responses to more than one vignette as the differences in the level of emotion might be larger between the scenes than between the conditions and we wanted the investigators to use the whole scale from 1 to 7 for each vignette.

Out of 253 valid questionnaires, in 246 there was a valid answer for the Milk scene, in 247 for the Money vignette and in 246 for the Moving Out scenario (not the same 246 than for Milk). To investigate for an effect of Condition, we performed a two-way repeated measures ANOVA on the factors Condition and Scene. We found no main effect of Condition, but a main effect of Scene ($F(2, 460)=14.805$, $p\approx 0$): The level of emotion was rated as significantly lower in the Milk scenario than for the two other scenes, which is in

line with previous results that the level of anger was rated as being the lowest in the Milk scene. There were interaction effects between Condition and Scene ($F(10, 460)=2.1517$, $p<0.05$), due to the fact that the level of emotion in the RH condition of the Milk scene was rated as being extremely low.

Who is mainly responsible for the argument? For each scene we performed a log-linear analysis of frequency tables with the variables Condition and Responsibility (three categories: woman, man, or no answer/both). No effect of Condition could be detected for the scenes Milk and Moving Out. For the Money vignette, there was an effect of Condition with $p<0.05$. It is mainly due to the fact that the man and the woman were rated as equally responsible by the participants of the NF condition. In all other conditions the woman was rated as responsible more often than the man (see Figure 4.13).

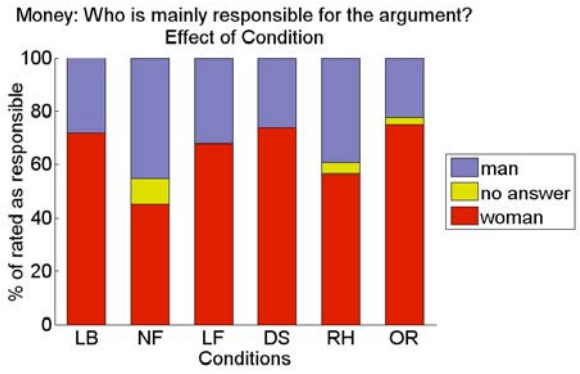


Figure 4.13: Results for the question "Who is mainly responsible for the argument?": There was an effect of Condition for the Money scene. The man and the woman were rated as equally responsible in the No Face (NF) condition, whereas the woman was rated as responsible more often in all other conditions.

What were they arguing about? The available space encouraged the participants to limit their answer to a few words. For each vignette, 7–8 categories were formed and each answer was allocated to exactly one category. When no answer was given and for answers as "don't know", the category "no answer" was used. To satisfy the exhaustiveness of nominal data scales, the category "other" was used if any of the answers would not fit into any other category. The resulting sets were tested for an effect of Condition with a log-linear analysis of frequency tables. For all three vignettes, there were effects of Category but no effect of Condition.

Please describe their physical interaction We asked this question as physical contacts seemed to be important in the interpretation of a scenario according to our experiment on the effects of desynchronization. Nevertheless, the answers were not very useful, mainly because the question was interpreted in very different ways. Some participants wrote about the story, others described the mood of the scene, others again depicted the physical motions of the characters.

Interpretative short free-form questions For each vignette, participants had to answer two questions that asked them to interpret a concrete part of what was happening in the scene. The questions were selected in such a way that the answer was suggested within the dialogue of the scene but a little space for interpretation remained. For example, at the end of the Milk vignette the man says: "Fine, I'll go to the store", which suggests an answer to the question "What does the man do after the scene?", while the participant can still choose to believe the man or not.

The response format was one line of free-form text. To analyse the responses, we chose a coding frame for each question and tested the resulting sets for an effect of condition with a log-linear analysis of frequency tables. The category "no answer" was included in the analysis.

When evaluating the frequency tables for "What is the woman trying to do when the scene starts?" and "What does the man do after the scene?" for the Milk vignette, "What does the woman try to prevent the man from doing?" for the Money scene, and "Why is the man there?" and "What does the woman do next?" for the Moving Out scene, we found a significant effect of Category ($p \approx 0$) each time, but no significant effect of Condition.

The only exception was the first question for the Money scenario: "What did the man do just before the scene?". Each answer was allocated to one of the categories: "went to mailbox", "checked credit card bill", "found a receipt", "get milk", "no answer" or "don't know", and "other". We found a significant effect of Category ($p \approx 0$) as well as a significant effect of Condition with $p < 0.05$. Closer examination showed that the results for the Original (OR) and the No Face (NF) conditions are distinct from the other conditions. More specifically, the percentage of answers in the category "no answer" was considerably higher and the percentage of answers in the category "went to mailbox" was lower for the conditions NF and OR, whereas the percentage of answers in all other categories remained similar throughout all conditions. That means that, similar to the question related to the

attention to details, exactly the two conditions in which the order of the scenes was changed have different answers. The Money scene was shown first for NF, second for OR and third for all others. We can therefore not exclude the possibility that the effect is entirely due to ordering effects.

Putting the three movies in what you believe to be the right chronological order, please tell the full story surrounding the events depicted. Based on the answers, we coded the order of the movies. When a vignette was mentioned to be first and another one second, the third one was completed, assuming that the participant ran out of time to write down the complete answer. The number of answers that differed from "Milk – Money – Moving Out" was very low. Therefore, we have to be careful how to interpret any results based on this data. The log-linear analysis of frequency tables with the variables Condition and Order showed no effect.

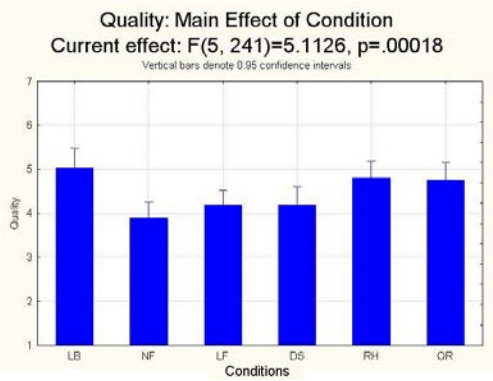


Figure 4.14: Results for the question "How would you rate the quality of the animation from a technical point of view?": There is a main effect of Condition for the quality ratings, the significant differences between means being LB>NF/LF/DS, OR>NF/DS, and RH>NF.

Quality

To analyse hypothesis H4, i.e. that degradations in character animation have an effect on the perceived quality of an animated scene, we analysed the quality ratings. We performed a one-way ANOVA on the factor Condition. We found a main effect of Condition with $p < 0.01$. The mean ratings are graphed in Figure 4.14. The post-hoc Newman-Keuls test showed the following significant differences:

- the quality of the Low Body (LB) condition is rated to be significantly higher than the quality of the conditions No Face (NF), Low Face (LF), or Desynchronized (DS)

- the quality of the Original (OR) condition is rated to be significantly higher than the quality of the conditions NF and DS
- the quality of the Rigid Hair (RH) condition is rated to be significantly higher than the quality of the condition NF

There was no effect between the ratings of the conditions LB, OR, and RH. Based on those results *we support hypothesis H4*: Degradations in character animation have an effect on the perceived quality of an animated scene.

4.2.3 Eye-Tracker Experiment Results

Contrary to the Group Experiment, where participants were sitting in lecture rooms, the Eye-Tracker Experiment was run in a small room with only the examiner and the participant.

In the next paragraphs, we give an overview of our participants and describe the results of the questionnaire and of the eye-tracker data.

Data overview

The number of participants with valid questionnaires for each condition are listed in Table 4.7. Only four questionnaires were discarded (one OR, one NF, and two RH conditions). The methods to analyse the data were the same as in the Group Experiment.

condition	AO	LB	NF	LF	DS	RH	OR	total
f	11	4	6	2	3	3	6	35
m	14	8	9	4	4	3	12	54
total	25	12	15	6	7	6	18	89
% f	44	33.3	40	33.3	42.9	50	33.3	39.3

Table 4.7: . Number and distribution of male and female participants with valid questionnaires for each condition of the Eye-Tracker Experiment in the degradation effect study.

To analyse the gaze data, we had to exclude numerous participants, due to calibration errors or other technical issues. Table 4.8 shows the number of valid participants for each scene and each condition. The numbers vary for each condition as the eye-tracker was recalibrated before viewing each animation. This was necessary because the participants might have changed their position relative to the screen when answering the questions between the scenes, which can decalibrate the gaze measurings.

vignette	LB	NF	LF	DS	RH	OR
Milk	6	8	6	6	7	8
Money	3	8	6	6	7	7
Moving Out	6	8	5	5	7	7

Table 4.8: Number of valid gaze recordings in each condition of the Eye-Tracker Experiment in the degradation effect study.

Questionnaire with Audio Only condition

The results for the questionnaire from the Eye-Tracker Experiment were very similar to those from the Group Experiment. The main effects of Scene were confirmed for the questions related to Angriness, Sympathy and Plausibility, as were the main effects of Character for the questions on Angriness and Sympathy and the interaction effects between Scene and Character for Angriness, Sympathy and Plausibility. No interaction effects with Condition were found, which might point out that those effects were due to differences of backgrounds between the groups. None of the questions related to the attention to details showed significant results, which confirms our assumption that those results were created by ordering effects in the Group Experiment.

For the question asking which character is responsible for the argument, contrary to the Group Experiment, we found no significant differences between the conditions for the Money vignette. However, for the Milk vignette, we found significant differences between the conditions ($p<0.05$, see Figure 4.15). The virtual woman was rated as being responsible for the argument the least in the LB condition, followed by the AO and NF conditions, and the most often in the RH and OR conditions. It seems that the more realistic the character, the more often the woman was rated as being responsible for the arguments.

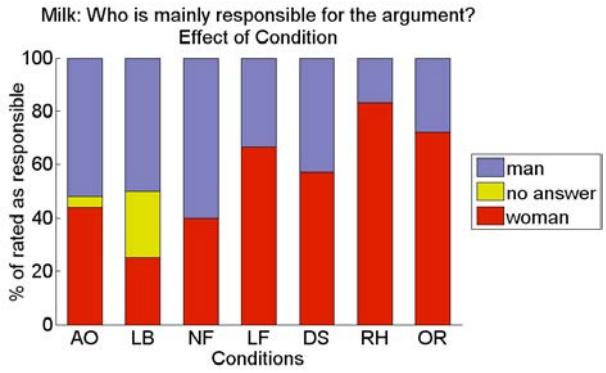


Figure 4.15: Eye-Tracker Experiment, results for the question "Who is mainly responsible for the argument?": There was an effect of Condition for the Milk scene.

Surprisingly, we did not find a significant effect for Quality, although the trend seems to be that the Audio Only (AO) condition was rated as having the lowest quality on average, followed by the LB and the NF condition.

Eye-Tracker

We know from previous research, such as [MLH*09], that people tend to gaze at characters' faces. Figure 4.16 shows the heat maps of the average visual attention for each vignette and condition. The areas that received the most attention from the subjects are red and yellow, scaling down to the blue and finally black areas that received the least attention. The location of the heads of both characters are marked by red or yellow regions in all conditions and vignettes because, as expected, they received the majority of the attention. We can see in the heat maps that, in general, people gaze more at the woman's face than at the man's face and that faces are looked at less in the Low Body condition than in the other conditions.

Therefore, we examine the regions of the virtual humans' heads more closely for each vignette and choose them as *areas of interest* (AOI). We measure and evaluate how long the gaze of the participants is within each AOI for each condition. To this end we use the *observation length* (OL), which is defined as follows: "Observations start when a fixation in an AOI starts and end when there is a fixation outside of the AOI. The observation length is therefore either equal to the fixation length or slightly higher as there can be small gaps between fixations without having a fixation outside of the AOI to stop the observation." [Tob08]. To evaluate the differences in OL across the conditions, we ran a repeated measures ANOVA with the dependent variable OL of the characters' heads AOI and the independent variable Condition for each scene. We did not include the AO condition in the evaluation of the eye motions, as there were no virtual characters in the scene and therefore significant differences with that condition would not give us any further insights.

As a general result, we found that participants looked significantly less at the faces in the LB condition, which is not surprising as the head is only represented by a deformed sphere. Furthermore, participants looked more at the virtual woman's face than at the man's face in all three scenes. This can be explained by the fact that she faces the camera whereas the man's face can not be seen well.

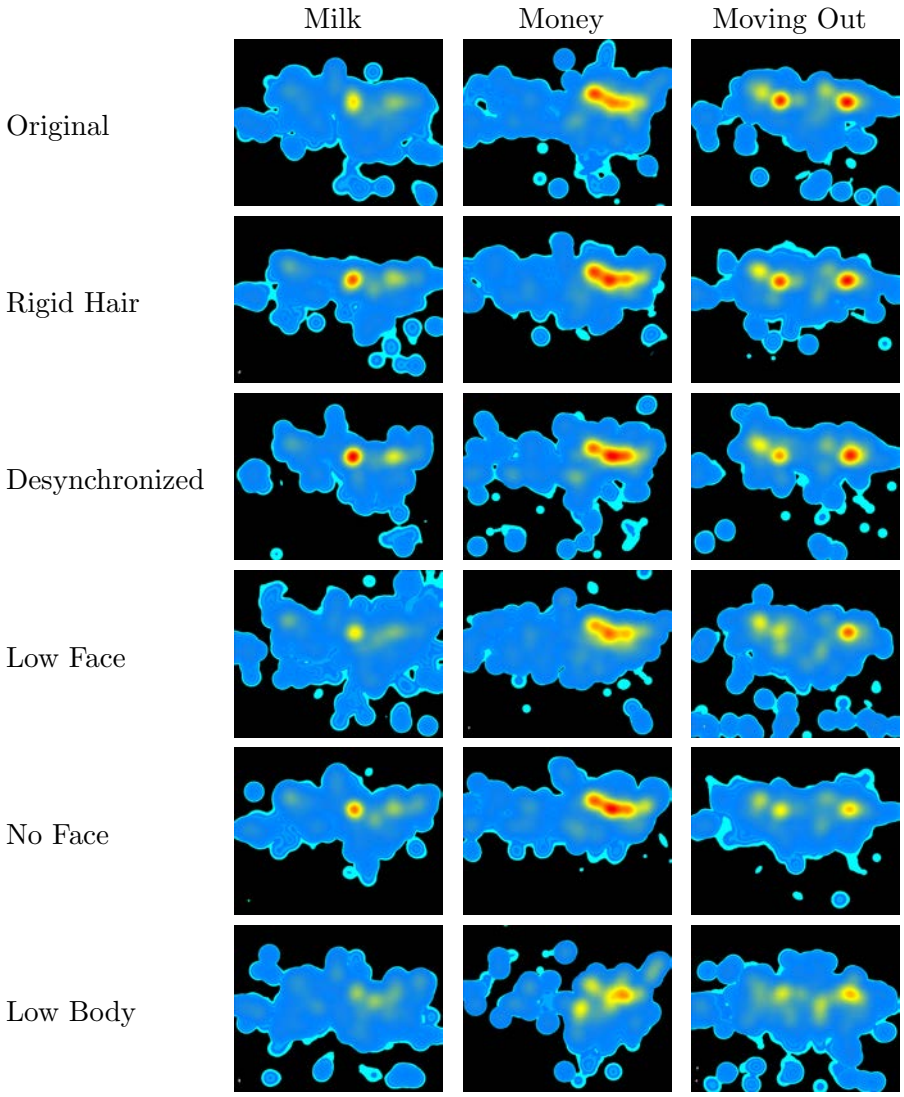


Figure 4.16: Heat maps for the different conditions and vignettes. The areas that have been gazed at most are displayed in red, followed by yellow and blue, and finally the areas that have barely been looked at are shown in black.

Milk In the Milk vignette, the female character moves her position considerably during the scene. She first stands briefly in one position with her head in the green area in Figure 4.17, then she moves, so that her head is in the red area. As the amount of time spent in the first position was very short, we only include the red area in our analysis. The position of the man's head is represented by the blue area.

As we analysed the observation lengths for the relevant areas of interest, we found a main effect of Condition with $p<0.005$ due to the fact that the faces were watched less in the Low Body (LB) condition. Furthermore, there were interaction effects between the Condition and the AOI. A Newman-Keuls test showed the following significant differences (see Figure 4.18):

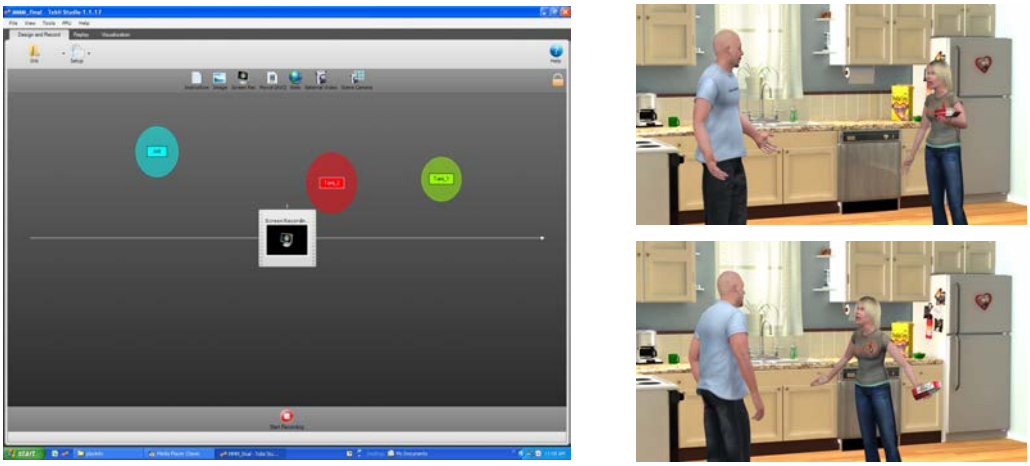


Figure 4.17: Areas of interest for the Milk vignette. The man's head position is marked in light blue. The woman's head position is in green for the beginning and in red for the main part of the scene. Because the time when the head is in the green area is very short, we only analyse data from the red area. Two frames from the Milk vignette show the locations of the characters.

- the AOI were gazed at significantly less in the LB condition than in the No Face (NF), Desynchronized (DS), Rigid Hair (RH), and Original (OR) conditions,
- there are significant differences between the observation lengths of the conditions DS and Low Face (LF), DS and OR, RH and LF, and RH and OR with the observations time being longer in RH and DS than in LF and OR.

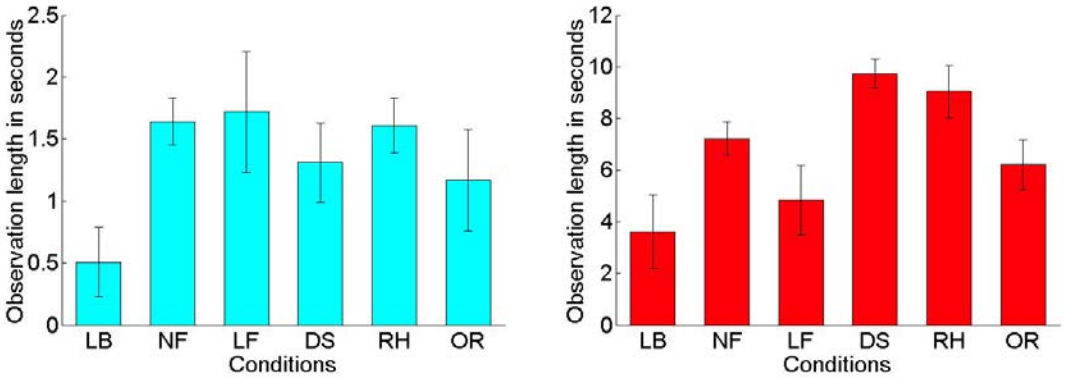


Figure 4.18: Milk: Observation length in seconds for each condition for the areas of interest corresponding to the male head (light blue area in Figure 4.17) and the female head (red area).

Money In the Money vignette, the two characters stay in the same positions for nearly the whole vignette. The AOIs, with the woman's head in yellow and the male's head in blue, are shown in Figure 4.19. Our analysis showed no main effect of Condition or interaction effects. Figure 4.20 graphs the observation lengths in each condition.

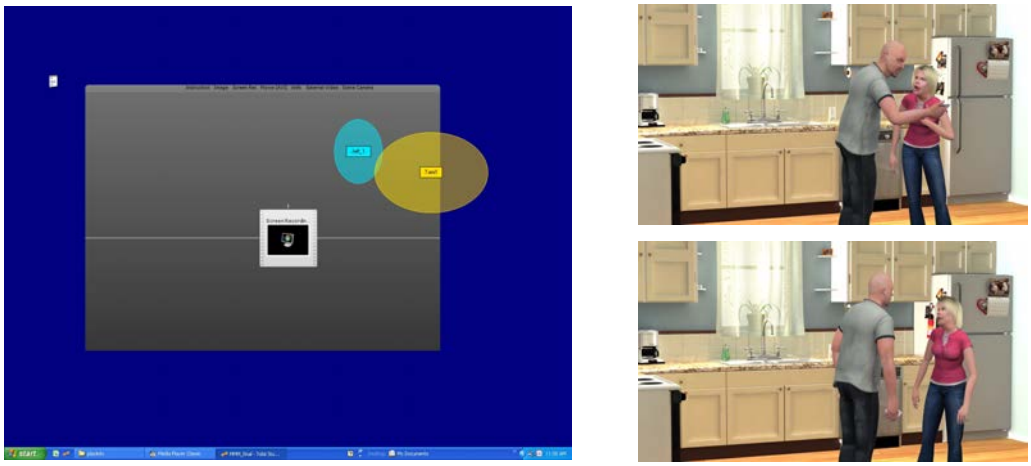


Figure 4.19: Areas of interest for the Money vignette. The man's head position is marked in light blue. The woman's head position is in yellow. Two frames from the Money vignette show the locations of the characters accordingly.

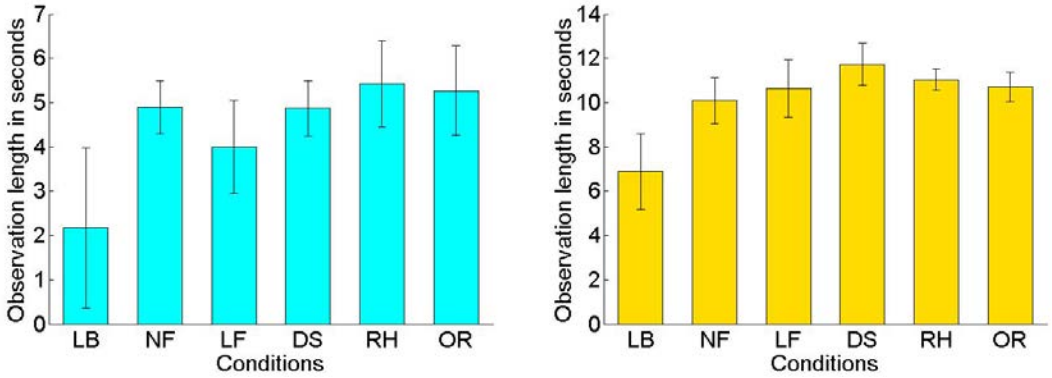


Figure 4.20: Money: Observation length in seconds for each condition for the areas of interest corresponding to the male head (light blue area in Figure 4.19) and the female head (yellow area).

Moving Out As the characters change their positions in the Moving Out scene, we split the scene in two parts that we analysed separately. The AOIs for both parts are depicted in Figure 4.21: dark blue is the man's face for Part 1, light blue the woman's face for Part 1, red is the man's face for Part 2, and green the woman's face for Part 2.

For Part 1, we found a main effect of Condition with $p<0.005$, due to the fact that the faces were watched less in the LB condition (significantly less than all but the LF condition). Furthermore, a Newman-Keuls test showed the following significant differences for the male character (see Figure 4.22):

- participants gazed less at the heads in the LB condition than in the RH and OR conditions
- the observations lengths were shorter for LF than for OR

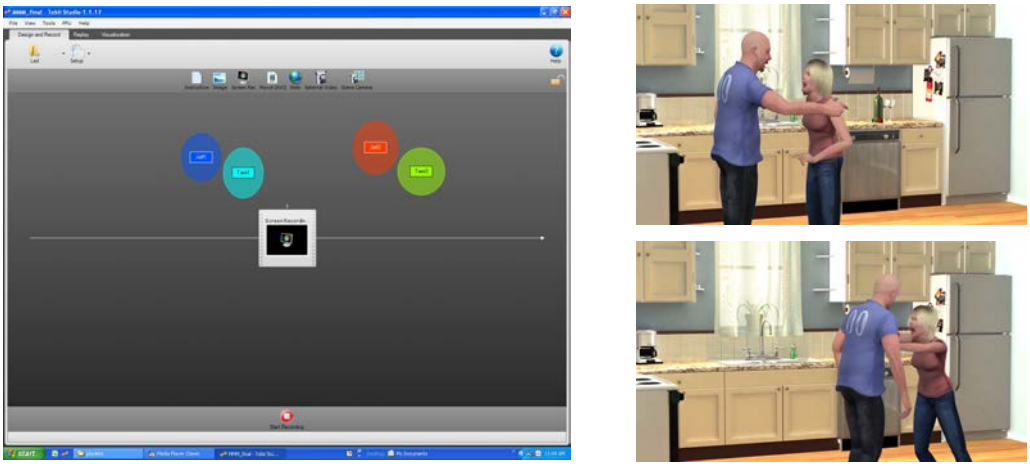


Figure 4.21: Areas of interest for the Moving Out vignette. The man's head position is marked in dark blue and red for part 1 and 2, respectively. The woman's head position is marked in light blue for part 1 and green for part 2. Two frames from the Moving Out vignette show the locations of the characters accordingly.

For Part 2, there is a main effect of Condition with $p<0.001$, again based on the fact that the faces were gazed at less in the LB condition (significantly lower than all conditions). The graphs are shown in Figure 4.23.

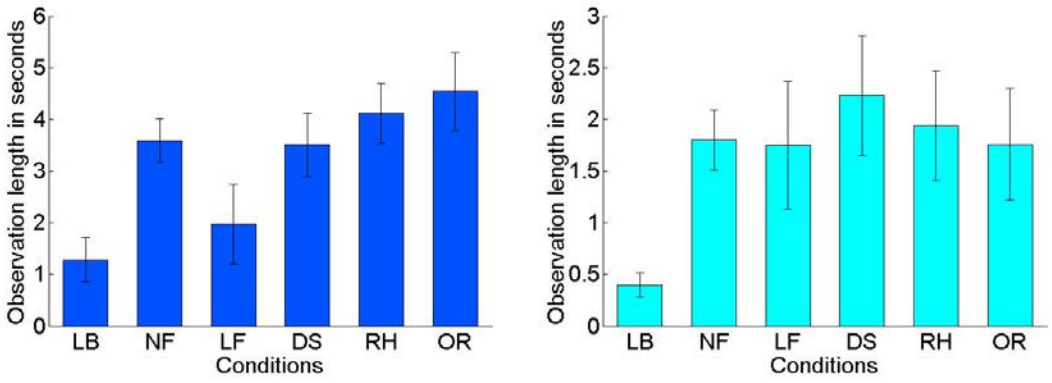


Figure 4.22: Moving Out, Part 1: Observation length in seconds for each condition for the areas of interest corresponding to the male head (dark blue area in Figure 4.21) and the female head (light blue area).

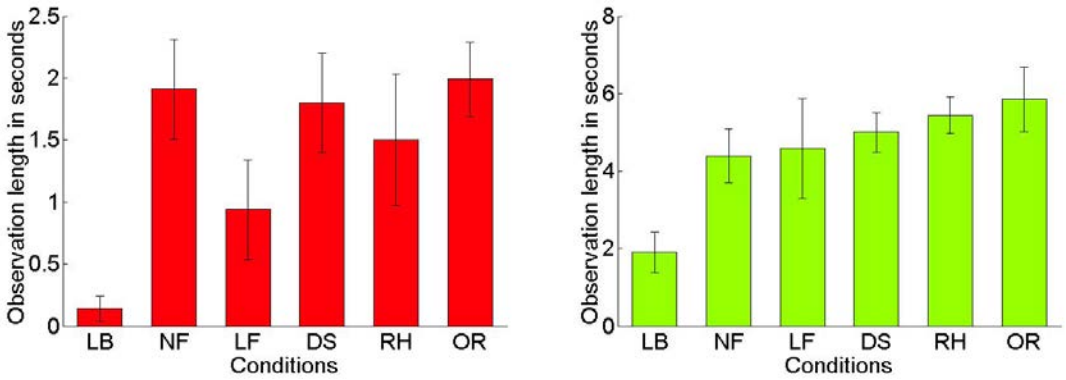


Figure 4.23: Moving Out, Part 2: Observation length in seconds for each condition for the areas of interest corresponding to the male head (red area in Figure 4.21) and the female head (green area).

4.2.4 Discussion

We discuss our results for the Group Experiment and the Eye-Tracker Experiment in detail within the next sections.

Group Experiment Discussion

One of our aims was to provide a set of stimuli and experimental methodology to investigate the effect of degraded motion. In order to do this, we wanted to create stimuli with distinct and increasing levels of emotion. The results of our analysis for the ratings in response to the question "How angry would you rate the characters in this animation?" show that we indeed succeeded in this aim. We showed that all three vignettes were rated as significantly different from each other, with the highest ratings for Moving Out, followed by Money and then Milk.

Within the ratings for the questions about angriness, appropriateness, and sympathy, we found several interaction effects involving the Condition: an interaction effect of Condition and Scene for angriness and interaction effects of Condition and Character for appropriateness and sympathy. It is notable that the Desynchronized condition is involved in the explanation of each of those interaction effects. This supports our results from the Synchronization Errors Experiment that errors in synchronization can have affect the perceived emotional content of a scene. Nevertheless, the differences are not as big as we expected and could also result from differences in the backgrounds of the participant groups. Another possible explanation for the fact that not all our differences reached significance is

that the audio communicated much of the emotional content of the vignettes and overrode some of the more subtle effects in the information communicated by the animation itself.

Evaluation of the Quality rating of the Group Experiment showed that the ratings for the Low Body (LB) condition with its very simple character model were as high as the ratings for the Original (OR) and Rigid Hair (RH) conditions, and significantly higher than the ratings for the No Face, Low Face, and Desynchronized conditions. This suggests that these simple models (LB), which are far away from a realistic human representation, could be on one side of the uncanny valley, whereas the characters with their original unmodified face and body motions (OR and RH) are on the other side. Low quality facial and body motion caused the animations in the No Face, Low Face, and Desynchronized conditions to drop into the uncanny valley. If we sort the conditions from the most stylized one (LB) to the one closest to a real human (OR), we could interpret the resulting shape in the graph (see Figure 4.14) as resembling the uncanny valley graph hypothesized by Mori.

The interactions between Character and Gender show that our scenarios might be perceived differently by men and women. Due to the constraints of our experimental design, the proportion of female participants ranged from 21.9% to 50.0% in the different groups. An even distribution of women across all conditions would have been preferable. Furthermore, our participants had a variety of educational backgrounds with the participants who had a similar background in the same group. As this was a between-groups design, we could not control the influence of this variable. Randomized assignment to the experimental conditions would have been desirable. It would be interesting to explore this effect in future experiments. If participants' backgrounds are not to be used as a dependent variable, differences between groups should be avoided.

The evaluation of two of the factual questions and of one of the interpretative short free-form questions showed significant effects of Condition. However, as described in Section 4.2.2, all three results could be the consequence of ordering effects. Future research could evaluate to what extent these effects are due to design decisions, such as the order of stimuli and the type of questions asked, which was not controlled for in our experiments. Otherwise, in this type of experiments, it is advisable to either present vignettes in exactly the same order to all participants or, if possible, to randomize the order for every single participant. The second option has the additional advantage that potential ordering effects in other answers, such as the ratings of the characters' angriness are excluded.

Eye-Tracker Experiment Discussion

The results from the evaluation of the questionnaire in the Eye-Tracker Experiment confirmed some of our results from the evaluation of the questionnaire in the Group Experiment, such as the different levels of emotional content in the three vignettes. It failed to confirm the interaction effects involving the Condition and differences in the attention to details, which might indicate that those effects arose from group effects and ordering effects, respectively. However, a different explanation could be a lack of statistical power given the relatively small number of participants in the Eye-Tracker Experiment.

The most interesting finding of this study was that there was no significant difference between the Audio Only (AO) condition and all other conditions. This result is counter-intuitive, as one would expect that the perception of the angriness of a character would change when neither body language nor facial expressions information is available. Apparently, sufficient information to perceive the emotional state of the characters was present in the sound alone.

We chose to include sound in our stimuli to increase the realism of our scenes because it is very unlikely that a modern movie would not have any audio. Nevertheless, we conclude that the audio was remarkably intense in our stimuli. In particular, participants' feedback to the question "What made you score the animation as high or low quality" indicates that the voice of the woman was very strident and annoying. For example, two participants viewing the Low Face condition in the Group Experiment wrote "not much movement in face or body (stiff, jerky movement). Voices gave the emotion, not animation" and "I didn't like the characters... There kinda scary looking... The woman's voice was always way more angry making me dislike her". These type of comments support our assumption.

It would be interesting to replicate our experiments without sound or with a different type of sound. The perception of multisensory stimuli with emotional content is not well understood, and the study of individual senses and elements in isolation does not suffice when trying to understand the complex processes involved. More experiments are needed to investigate the processes that affect people's perception of a movie. Therefore, we support the use of emotional and multisensory content in experimental stimuli for computer graphics.

Summary and Future Work

The Group and Eye-Tracker Experiments resulted in a multitude of results. We could not find enough evidence to clearly support hypotheses H1-H3. This means that we did not find support that a variety of degradations in character animation alter the emotional content, the attention to details, or the interpretation of a scene. However, this does not mean that degradations never alter those variables. For example, we showed that some type of degradations, such as errors in synchronizations, might change the emotional content. Probably this effect was outweighed by the predominant sound of our vignettes. Our result investigating the quality ratings is more definite: we found support for hypothesis H4, and showed that degradations in character animation can alter the perceived quality of a scene.

Our main findings can be summarized as follows:

- Three vignettes with different levels of emotional content were created.
- Degradations in character animation can alter the perceived quality of a scene.
- Errors in synchronization may alter the emotional content of a scene.
- Audio conveyed the main information of our vignettes.

Furthermore, we found several guidelines for future experiments, which would help to clarify and interpret differences in ratings:

- Randomized allocation of participants to conditions is necessary.
- Even distribution of male/female participants in every group is necessary.
- Order of stimuli needs to be randomized across participants.

In our study, we chose six conditions plus the Audio Only condition. Many other conditions might also have produced interesting results. For example, hand gestures are often credited with communicating significant information in speech and were found to have an effect on the interpretation of a scene in our study analysing the effect of synchronization errors (see Section 4.1). A motion capture skeleton with fewer degrees of freedom in the shoulders and back might also have been an interesting condition to evaluate.

We implement several of these findings in our next experiment.

4.3 Validation Experiment

We design a study that takes our results from the previous experiment into consideration[1]. We explore the relative importance of several anomalies using two methods. First, we use a questionnaire to determine the emotional response to the full-length vignettes Milk, Money, and Moving Out, with and without facial motion and audio (four conditions) in the *Vignette Experiment*. Second, we compare the performance of a virtual actor in short clips depicting a range of different facial and body anomalies in a 2AFC (two-alternative forced choice) task in the *Clip Experiment*. The clips are extracted from the vignettes. The selected anomalies explore two classes of conditions: common flaws in animation, such as the eyes not tracked, and disease conditions, such as tremor or Bell's Palsy.

Based on the results of our previous experiment, we study the interaction of sound with degradations in animations and therefore include vignettes with audio and without audio in our experiment. As each degradation in animation coupled with audio/no audio would add two more conditions to the Vignette Experiment, we focus on one of the most obvious degradations: no facial animation. For the Clip Experiment, the flaws in animation are selected to give practical advice on how to create compelling characters and with the disease conditions we investigated whether the repulsion experienced when watching characters that fall into the uncanny valley are caused by animations that resemble disease conditions as postulated by MacDorman [Mac05].

Our hypotheses are as follows:

- H1: Facial anomalies will have the most effect as McDonnell et al. [MLH*09] and others have shown, that people focus most on the faces of real and virtual humans.
- H2: Conditions relating to human diseases will have a similarly strong effect.
- H3: The presence or absence of audio will influence responses, as it is to be expected that much of the emotional content of a vignette is delivered through this medium.

4.3.1 Method

Similar to our previous experiments, we first describe our stimuli, participants, and procedure, before we present the results and discuss them.

[1]This study is shared work with Jessica Hodgins, Carnegie Mellon University and Disney Research, Pittsburgh and Carol O'Sullivan, Trinity College Dublin

Stimuli

To test the above hypotheses, we used the same three vignettes depicting an arguing couple than in the previous experiment. However, as our previous experiment indicated that the sound was very important in the impact on our participants, we re-recorded the dialogues with the same actors in a separate recording session to generate a less overwhelming version. For the Vignette Experiment four conditions were created for each vignette: with and without facial motion and audio (FullFaceSound, FullFaceNoSound, NoFaceSound, and NoFaceNoSound). For the NoFace conditions, we rendered the animations with no facial or eye movements, while for NoSound we turned off the audio.

To create the stimuli for the Clip Experiment, we divided the vignettes up into shorter snippets that still made sense, for example a full sentence, each between 3-6 seconds long. By modifying the original animation, we created six facial and three body anomalies that we applied to our female character. Multiple clips were selected for each condition. This set of stimuli was sufficient to allow random selection in the experiment and thus reduce repetition. The Clip Experiment conditions were as follows:

- Full (F): the original, unmodified animation.
- NoFace (NF): the face and eyes were completely stiff.
- NoEyes (NE): the face was animated but the eyes did not move (the character was still blinking).
- OffsetEyes (OE): one eye was rotated by a constant amount, giving an effect similar to amblyopia (or lazy eye).
- LowFace_Eyes (LF_E): simplified facial animation where the motion of the jaw was restricted to one degree of freedom (open/close for speech synchronization), the eyes were animated.
- LowFace_NoEyes (LF_NE): face as with LF_E, but the eyes were not animated
- HalfFace (HF): only half the face was animated – this condition was chosen to mimic Bell's palsy.
- NoArm (NA): one arm was not animated and was held stiffly in front of the character as if in a sling.
- NoisyArm (NyA): a tremor was added to the motion of the character's arm.

stimuli	condition	participants
vignettes	FullFaceSound	21 (13m, 8f)
	FullFaceNoSound	15 (10m, 5f)
	NoFaceSound	21 (11m, 10f)
	NoFaceNoSound	14 (10m, 4f)
total vignettes		69 (48m, 37f)
clips	FaceType	35 (18m, 17f)
	FaceType/Body	34 (24m, 10f)
	NoFace	16 (10m, 6f)
individuals (= total clips)		85 (52m, 33f)

Table 4.9: Number of participants in each condition of the Vignette and Clip Experiments in the validation study. Most participants took part in both experiments.

- NoPelvis (NP): only the top half of the body was animated by removing data from the root and legs. The motion of the two spine joints was smoothed to reduce artefacts caused by the lack of motion in the pelvis

Our goal in the design of these experiments is to create large effects that are certain to elicit differentiated responses. Therefore, we aim to generate clearly suprathreshold stimuli, which are well above the Just Noticeable Difference for the conditions tested. We did informal pretests with people who were familiar with the animations to tune the magnitude to what was definitely noticeable without being too extreme. We are confident that all stimuli were approximately equally noticeable when viewed directly by an expert.

Participants

The vignettes and clips were seen by small groups of participants to avoid the group effects that might have been present in our previous experiment (see Section 4.2). A total of 28 small groups participated, ranging in size from one to six people. This enabled a randomized allocation to conditions and a randomized order of clips for each group. We recruited participants by advertising on university e-mail lists, posters, and fliers. This recruitment effort resulted in a varied participant pool drawn from a range of different disciplines and backgrounds: 85 individuals (52m, 33f) ranging in age from 18-45, with 76% of them reporting passing knowledge or less with 3D computer graphics. All gave their informed consent. Out of the 85 participants, 69 did both the Vignette and the Clip Experiment; the remaining 16 only participated in the Clip Experiment at the end. The numbers of participants that were subjected to the different conditions are listed in Table 4.9.

Procedure

At the start of the experiment, the participants filled out some general information such as gender and age. Those who participated in both experiments first watched all three vignettes in the same order (Milk, Money, Moving Out) displayed in one of the four conditions. After each vignette, we displayed questions on the screen. These questions required them to rate on a 5-point scale the anger levels of both characters, how much they empathized with them, how much sympathy they had for them, and how justified their reactions were. They were also asked to select who was mainly responsible for the argument on a 5-point scale (woman only, woman mostly, both, man mostly, man only). At the end of the Vignette Experiment, they were asked to list objects and differences they had noticed and to write five short sentences about the events.

Then the Clip Experiment started and they watched a random ordering of the clips after the scenario was set as follows:

> We have captured the voice and motions of the female actor and now want to know which virtual characteristics are best at conveying a compelling performance. Therefore, you are "auditioning" a virtual actor. You will be asked after viewing each pair to indicate your answer to the following question: "In which clip was the virtual actor more convincing?"

The participants indicated their choice for each pair of clips on an answer sheet. Asking to choose the more convincing actress is at a higher level than simply looking explicitly for anomalies, for example, by asking in which clip they saw an animation artefact. Each group viewed a random selection of four clips extracted from different parts of the vignettes so that participants would not get too accustomed to a specific sentence. The order of clips varied throughout the small groups to avoid any ordering effects. They saw every combination of pairs of conditions within one of three sets:

- FaceType: the six different conditions related to the face and the unmodified animation, i.e. the conditions Full, NoEyes, OffsetEyes, LowFace_Eyes, LowFace_NoEyes, and HalfFace,
- FaceType/Body: the five disease conditions and the unmodified animation, i.e. the conditions Full, OffsetEyes, HalfFace, NoArm, NoisyArm, and NoPelvis,
- NoFace: the three most obvious artefacts as determined by the first two sets of

conditions and the NoFace and Full conditions, i.e. the conditions Full, NoFace, OffsetEyes, HalfFace, and NoArm.

The conditions for the Vignette and for the Clip Experiment were randomized independently in each experiment so that the small groups saw different combinations of conditions, with the vignettes always shown before the clips.

The Vignette and the Clip Experiment took approximately 20 minutes each. The experiments were all held in the same room, where all videos were projected onto a 150 x 80 cm screen at a distance of 2-2.5m. We asked participants not to make audible responses to the animations and they were not permitted to discuss their opinions with each other during the study. An experimenter was present at all times to ensure compliance and to control if the answer sheets were filled out correctly. Furthermore, the participants were given as much time as they needed and they were rewarded with a 5 euro book voucher for each experiment (resulting in 10 euro if they participated in both, the Vignette and Clip Experiment). This way, we avoided missing answers, lack of motivation, and errors in filling out questionnaires as they happened in our previous experiment.

4.3.2 Results

The Vignette Experiment was analysed with a three-way, repeated measures ANOVA with the factors:

- FaceType, between-subjects variable, 2 values (FullFace, NoFace),
- Audio, between-subjects variable, 2 values (Sound, NoSound), and
- Scene, within-subjects variable, 3 values (Milk, Money, Moving Out).

We conducted post-hoc Newman-Keuls tests for pairwise comparison of means. The most significant results are shown in Figure 4.24 and a summary of significant effects (i.e. with $p < 0.05$) is shown in Table 4.10.

The most significant results were found for the woman, which was expected as her face was visible throughout the three vignettes. There was a significant positive correlation between the ratings for the sympathy and justified ratings (Spearman's $\rho = 0.75$, $p < 0.05$). Therefore, we do not include these results, or those relating to questions about the man, as they provide no added insights. There was also a significant correlation between sympathy

Question	Effect	Probability	Reason
Woman Angry	Audio∗FaceType	$F(1,65) = 9.8401$	FullFaceSound < FullFaceNoSound NoFaceNoSound < FullFaceNoSound
	Scene	$F(2,130) = 61.229$	Money < Milk < Moving Out
	Scene∗FaceType	$F(2,130) = 7.7607$	Money NoFace < Money FullFace < all others all others < Moving Out FullFace < Moving Out NoFace
Woman Sympathy	FaceType	$F(1,65) = 5.1201$	FullFace<NoFace
	Scene	$F(2,130) = 16.923$	Money < Milk < Moving Out
	Scene∗Audio	$F(2,130) = 6.0725$	Money Sound < all
	Scene∗FaceType	$F(2,130) = 4.1188$	all < Moving Out NoFace
Man Angry	Scene	$F(2,130) = 44.479$	Milk < Moving Out < Money
Man Sympathy	Scene	$F(2,130) = 6.4253$	all < Money

Table 4.10: Significant results of the Vignette Experiment in the validation study ($p < 0.05$ in all cases).

FaceType *Main Effect:* $F(5,170) = 66.57$	**FaceType/Body** *Main Effect:* $F(5,150) = 33.24$	**NoFace** *Main Effect:* $F(4,60) = 131.38$
Full > LowFace_NoEyes, LowFace_Eyes Full > OffsetEyes, HalfFace NoEyes > LowFace_NoEyes, LowFace_Eyes NoEyes > OffsetEyes, HalfFace LowFace_NoEyes > OffsetEyes, HalfFace LowFace_Eyes > OffsetEyes, HalfFace OffsetEyes > HalfFace	Full > NoArm, OffsetEyes, Half Face NoPelvis > NoArm, OffsetEyes, HalfFace NoisyArm > OffsetEyes, HalfFace NoArm > OffsetEyes, HalfFace	All significantly different

Table 4.11: Significant results of the Clip Experiment in the validation study ($p < 0.05$ in all cases).

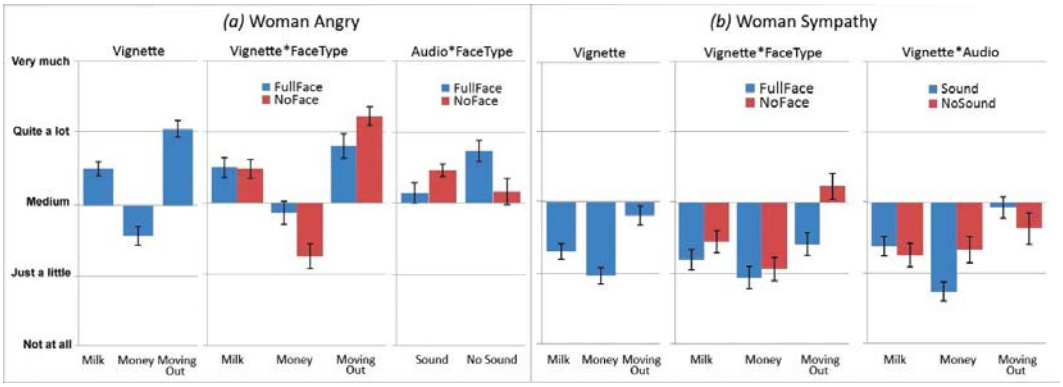


Figure 4.24: Most interesting results for the Vignette Experiment in the validation study.

and anger ratings, but this was much lower ($\rho = 0.26$, $p < 0.05$) and different effects were observed for this question.

As expected, a main effect of Scene was found for all questions. In particular, the Milk, Money, and Moving Out vignettes were found to elicit distinct levels of anger and sympathy ratings for the woman (see Figure 4.24 and Table 4.10). However, the order of the effect (Money < Milk < Moving Out) was different from the order in our previous experiment (Milk < Money < Moving Out), which means that the new recording of the voices had an impact on participants' responses.

An interaction effect between Audio and FaceType for the woman's angriness revealed that she was perceived as angriest in the FullFaceNoSound condition and much less so when either sound was present or the facial motions absent (see Figure 4.24 (a)). This result suggests that, while her facial animation did appear to express anger, the sound appears to have diluted the emotion, perhaps because we recorded the soundtrack separately. An interaction effect between Vignette and FaceType showed that she was perceived most angry in the Moving Out vignette with NoFace, and least angry in the Money vignette with NoFace. Maybe the lack of facial animation in Money reduced the overall level of expression, whereas in the Moving Out vignette, the lack of facial expression distracted less from the more physical body interactions between the couple. Or it may have been the case that her "stony" face simply came across as extremely angry in that vignette.

For the question on the woman's sympathy (see Figure 4.24 (b)), a main effect of FaceType is explained by the fact that the woman was found to be significantly less sympathetic when her face was animated (FullFace) than when it was not (NoFace) (not shown on graph). The lower sympathy rating with the facial animation actually implies a stronger

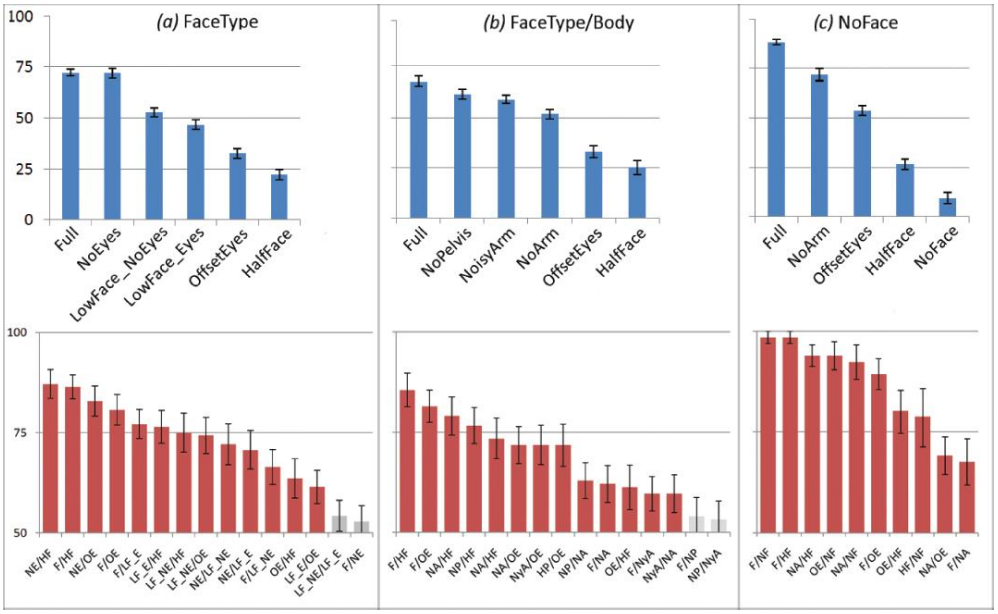


Figure 4.25: All results for the Clip Experiment in the validation study: overall (top) and pairwise (bottom) preferences. The y-axis displays the percentage of times the condition is preferred; error bars show one standard error.

negative reaction to the woman. An interaction effect between Scene and FaceType occurred because she was rated most sympathetic in the Moving Out vignette with NoFace. However, this result was closest to neutral and therefore implies a lack of expressivity. Therefore, the animated face results imply a stronger negative reaction to her, perhaps because her anger came across as too strong for the circumstances and hence people disliked her more. This was reflected in some of the comments from the FullFace conditions, such as "The woman seemed more aggressive and louder"; "The man was more explaining than arguing"; "Physically aggressive woman". The interaction effect between Scene and Audio was caused by the woman being least sympathetic in the Money vignette with Sound.

These results reinforce the common wisdom that facial animation is a key component of expression and careful attention should be paid to how the face is used to convey emotion. However, how the audio is matched or dubbed to the facial animation is also clearly key, as it interacted with the other factors in non-obvious ways.

The results for the Clip Experiment are shown in Figure 4.25. To compare the means of the overall preferences for the FaceType, FaceType/Body and NoFace sets of clip conditions (shown on the top), we carried out a single factor, repeated measures ANOVA for each. Post-hoc analysis was performed using Newman-Keuls tests and the significant effects are

reported in Table 4.11. For the pairwise comparisons (see Figure 4.25, bottom), single t-tests were carried out to determine if they were real preferences (i.e. significantly different from 50%). Non-significant effects are shown on the graph in grey.

As expected, the unmodified animation was preferred overall, except to NoEyes (see Figure 4.25 (a)). Probably the lack of eye movements was not more disturbing because we included eye-blinks. Apart from the completely stiff face, the two conditions related to facial diseases were found to be the worst as hypothesized, with OffsetEyes, the condition referring to amblyopia, being slightly but significantly more preferred to HalfFace, the animation mimicking Bell's palsy. The participants that compared these two conditions with body anomalies (FaceType/Body) found the facial anomalies significantly more disturbing than any of the quite significant body errors (see Figure 4.25 (b)). In fact, some of the participants reported not noticing the body errors at all. So, we can conclude that facial anomalies are more salient than errors related to the body. In the last condition set, where we tested the effect of NoFace to determine whether it really was the worst condition, we found that the NoFace condition was significantly less preferred, even to the worst of the other anomalies as shown in Figure 4.25 (c).

4.3.3 Discussion

We have developed a new type of experimental framework that did, as intended, elicit higher-level responses from participants to ecologically valid stimuli, similar to those that would be found in real applications. Our experiments now offer a starting point for future research to design new experiments and methods with the goal to run new experiments to provide further guidelines and insights for animators and the developers of animation systems. These promising approaches to investigating these issues are also applicable for evaluating other types of problems – not just motion capture.

In summary, in the Vignette Experiment we found that removing facial animation and/or sound did change the emotional content that was communicated to our participants. The woman's behaviour was found to be particularly affected by these four conditions, as she was the most visible in the scenes. Clearly, even the animations with no sound and no facial animation conveyed emotional content, but less effectively than when facial animation was included (making the woman more unsympathetic, for example).

The results from the Clip Experiment were unambiguous. Facial anomalies are more

disturbing than quite significant body motion errors, so this is where the most effort should be expended to achieve natural human motion. While we expected that facial anomalies would dominate over body motion errors, it was more significant than we had predicted. Furthermore, we could also derive some more nuanced guidelines regarding the significance of different types of facial anomalies relative to one another. An interesting observation is that, although the NoFace condition in the vignettes conveyed emotional information to the participants even when sound was absent, in the clips it was almost never preferred to any of the other conditions. However, in the written comments after viewing the vignettes, several participants referred to the absence of facial motion and expressed annoyance. This observation is worthy of further investigation, as are the effects of other types of emotion.

What do these results tell us about the uncanny valley theory? As we predicted, the two face disease conditions and the broken arm were preferred least, therefore supporting a hypothesis that these motions fell into the uncanny valley because they reminded people of illness or injury. On the other hand, the facial anomalies were much more disturbing than the body motion errors, which shows that the valley may be affected by attention. The frozen face was preferred least of all in the clips, perhaps because it resembles a corpse, yet the woman in the Moving Out vignette without facial animation was rated as being most angry. This could have been caused by her appearing “stony-faced”, which is an extreme form of anger. Alternatively, it may have been caused by focussing more attention on the quite aggressive body language of the couple, again pointing to a possible effect of attention. Eye-tracking studies would be very useful here to determine if this is indeed the case. Audio was also an important factor when watching the vignettes, as it interacted with both, the scene and the facial animation, in different ways. This effect was most evident for the Money vignette, where the absence of sound actually increased the perceived anger of the woman. All of these results point to a complex, multi-dimensional model of uncanniness, which is unlikely to be a valley but rather a parameterizable space.

We intentionally designed our stimuli to be suprathreshold. However, without calibrating the different magnitude changes we applied to our stimuli, it is difficult to rank order them in the perceptual effect that they have (i.e., one could argue that it might be that one change was ”bigger“ than another rather than that faces are more important than arms). Therefore, psychophysical studies to explore Just Noticeable Differences (JND) would provide further useful insights. We could, for example, run JND experiments, fit an Ogive function to the data, and then run our experiments on, say, the stimuli

at 75% perceptibility. Now that we have an effective framework for eliciting responses from participants, and some indicative results, we can explore how to probe around these extremes in even more targeted ways to expose greater detail about the responses and to appropriately model these data.

Other researchers have pointed out the importance of synchrony between different modalities [CVB01]. Nevertheless, the effects of the interactions of different modalities (e.g. voice intonation, body gestures, facial motion, or eye gaze) on viewers are still not well understood and need to be investigated more deeply. In our Vignette Experiment some participants seemed to suggest that there was a mismatch in emotion between the body and the voice. One participant, for example, wrote: "Both have unnaturally calm voices given their physical actions". Such a mismatch could lead to the perception of unfelt emotions.

We were able to run only a limited set of our conditions as we used a between-sujects design. It would be interesting to study the effects of other types of degradations. Further insights could also be gained by comparing our method to other types of metrics, including physiological responses such as heart rate, galvanic skin response or skin temperature [MIWFPB02].

Chapter 5

Hands

In Section 4.1 we found that finger motions are crucial in our understanding of a scene as participants' interpretation of a scenario was considerably altered by a 0.5 second error in synchronization of the finger motions. The great impact of finger motions is not surprising. Hand and finger motions are omnipresent in daily life. We use our hands naturally to punctuate our speech or to handle complex tools. In fact, as psycholinguistic research shows, hand movements and finger gestures play a substantial role in communicating information and meaning [McN92]. Given the importance of such motions in real life, they are also likely to affect our perception of virtual characters, especially when those characters are very human-like or interact with people. Therefore, in this chapter we focus on hand motions as a particular aspect of human motion.

Even though motion capture has become a widespread technology to animate virtual characters for movies and games, as discussed in Section 3.1, the subtle motions of the fingers are still complex to capture. Due to their small scale and frequent occlusions, optical motion capture of fingers requires careful planning, a larger number of cameras to cover the same space, and laborious manual post-processing. Therefore, finger movements are often animated manually or left out altogether. It is also possible to capture them separately from the body and to subsequently synchronize them, which might lead to visible artefacts [MZF06]. Little is known about the perception of such artefacts or about the effect of missing finger motions. The task of animating hands is also particularly complicated due to their elaborate structure of joints, tendons, and muscles leading to a large number of degrees of freedom (DOFs). However, from observation it can be seen that single rotations of specific joints are strongly correlated, meaning that they are performed

together in many situations. For example, the ring finger and the little finger are often bent at the same time. Furthermore, other joints, such as the distal interphalangeal joints, barely move most of the time, which adds even more potential to simplify the animation task.

In this chapter, we present our research related to finger motions. As previously mentioned, with the term "fingers", we refer to all five digits including the thumb. Creation or analysis of finger motions in this document includes all phalangeal joints and the carpometacarpal joint of the thumb, but excludes the wrist.

In Section 5.1, we evaluate the perceptibility and perceived quality of errors in finger motions. Many studies exist in the areas of human motion perception and hand motion generation as we showed in Chapter 2. However, to the best of our knowledge, the perception of hand motion for virtual characters has not been systematically analysed yet. Our study therefore aims to bridge the gap between research on finger animation and human motion perception. From previous research on human motion perception we know that people have impressive abilities to recognize and interpret detail in human body motion. In Section 4.1, we showed that this is the case for finger motions as well. In our experiment, participants described what happened in a scene depicting a man getting angry when operating a computer. Their interpretation of the story differed when the finger motions were desynchronized compared to when they were correct. The quality ratings, nevertheless, remained the same in both versions. This means that incorrect movements changed the understanding of a scenario, even without affecting the viewer's perceived quality. With this study in mind, we investigate two further questions related to the perception of finger motions: Firstly, what threshold of error can be detected when body motions and finger motions are not adequately aligned in time? Animations below this just noticeable error in synchronization would not need to be considered. Secondly, which type of animation method produces the best perceived quality? We establish a ranking between no motion at all, motion captured movements, keyframed animation, and randomly selected motion captured movements. For each question, we produced appropriate stimuli (see Figures 5.2 and 5.3) and ran a set of perceptual studies described in Sections 5.1.1 and 5.1.2.

Section 5.2 presents our approach to reduce the dimensionality of hand motions. We analyse correlations in hand motions and thereby identify ways to simplify the generation of hand movements while keeping individual joint rotations as a basis for the motions. Such methods allow a keyframe animator to only animate a few joints with the remaining

rotations being generated automatically. Hereby, we simplify the task of animating hands. Furthermore, this approach enables us to measure fewer joint rotations during motion capture sessions and thus to reduce the number of markers and the time needed for post-processing. Our approach uses Root Mean Square (RMS) deviations as a distance metric to analyse the correlations between different degrees of freedom of hands. Although other researchers have used distance metrics and dimensionality reduction techniques to handle motion capture data [GM08, BSP*04], we apply them to a novel field and purpose. In previous studies, principal component analysis (PCA) has been used most often as a dimensionality reduction technique. PCA results in a new set of basis rotations, each in general consisting of rotations of all joints. That means that changing the influence of one principal component (or eigenvector or new basis vector) would move a multitude of joints and the rotation of a single joint would be influenced by several components. Our goal is to reduce dimensions while keeping the basic rotations of the hand separate. With our approach, each rotation value depends on one other joint at most, so that we have an intuitive system that an animator can work with. Using the findings of our analyses, we develop a method that allows us to simplify the task of animating or capturing hands with minimal error in the resulting motions.

Finally, we use our acquired knowledge on hand perception and animation in Section 5.3. The crowd simulation system *Metropolis* is able to steer several thousand pedestrians walking through a virtual version of Trinity College Dublin and its surroundings. Groups of standing and talking people can be generated as well. Nevertheless, none of those animations includes hand motions as the capturing of hand motions would have been too complex. We implemented the necessary pipeline to make it possible to add previously generated finger motions to these animations in the future.

5.1 Perception of Hand Animation

5.1.1 Just Noticeable Errors

In our first experiment, we study the amount of synchronization error detectable by the viewer. We focus specifically on this type of error as slight desynchronizations are hard to avoid when body and hand motions are captured separately. Our experimental design follows the method of constant stimuli, which is well established in psychophysics. We

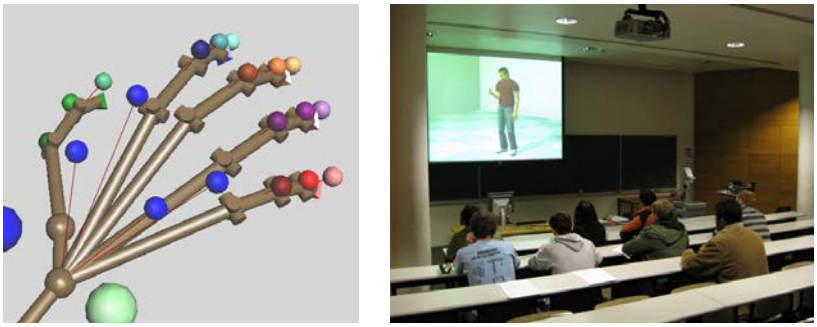
ask participants to categorize short gestures as being simply motion captured or modified, showing unmodified and modified stimuli in equal number and in random order. We found that even synchronization errors as small as 0.1 seconds caused a significant difference in the scores. However, this result cannot be generalized as the amount of imperceptible error depends on the performed gesture.

In the remainder of this section, we describe our methods (stimuli, participants, procedure) and analyse and discuss the results that we obtain.

Stimuli

We generated four short sequences involving finger motions: counting, drinking from a bottle, pointing, and snapping. We chose the gestures to be as varied as possible. *Point* is a classical gesture with three phases (preparation, stroke, and retraction). *Count* and *Snap* have cyclical repetitions in their main part. *Drink* has a more complex structure. It involves opening a bottle and then drinking from it. Snap and Drink include interactions with an object or self-interactions of the fingers. All gestures are self-explanatory, very common, and easy to recognize.

We captured the body and finger motions of a male performer simultaneously, using 19 markers attached on each hand and 48 on the body, with a 12 camera Vicon optical system (see Section 3.1 for more details on the capture process). To prevent the participants from becoming too familiar with the motions, three versions of each action were recorded, resulting in 12 distinct original gestures. The movements of the skeleton (see Figure 5.1 (a)) were then calculated and displayed on a virtual human (see Figure 5.2). For each version and action, four levels of modification were generated. The keyframes of both hands' finger joints – including all phalangeal joints and the carpometacarpal joint of the thumb, but excluding the wrist – were moved in time. Based on a pilot experiment, we delayed the finger motions by 0.1, 0.2, 0.3, and 0.4 seconds (3, 6, 9, and 12 frames). Modified and unmodified motions were counterbalanced as, in the pilot experiment, this seemed to encourage participants to try harder to identify modifications, therefore increasing sensitivity. As a result, every participant watched 4 (actions) x 4 (modifications) x 3 (versions) x 2 (to counterbalance) = 96 motion clips (48 modified and 48 non-modified ones).



(a) (b)

Figure 5.1: (a) Hand skeleton calculated based on the marker positions. (b) Group participating in the experiment investigating just noticeable errors.

Participants

Twenty-two participants (4f, 18m, aged 18-35) viewed these clips in two groups and in different random orders. Eight people were present at the first screening and 14 people at the second. The participants were mostly students and were recruited through e-mail lists and advertisements at Trinity College Dublin. Each participant received a 5 euro book token.

Procedure

The motions were projected on a large screen in a lecture theatre as in Figure 5.1(b). At the beginning of the experiment, we explained what motion capture is by showing a video of an actor in a motion capture suit and the corresponding virtual character performing the same motion. We furthermore asked them for basic information, such as their field of work, their gender, and their exposure to virtual characters or computer graphics. Then, for each clip, participants were asked: *Is this motion MODIFIED or NOT MODIFIED*? Between each animation a black screen was displayed for 5 seconds, during which participants noted their answer. After 3 seconds, they were alerted to the next stimulus by a sound. We added two clips at the beginning of the experiments, which were not included in the evaluation – one being modified, one unmodified – to allow the participants to get familiar with the procedure. The experiment took about 25 minutes in total.

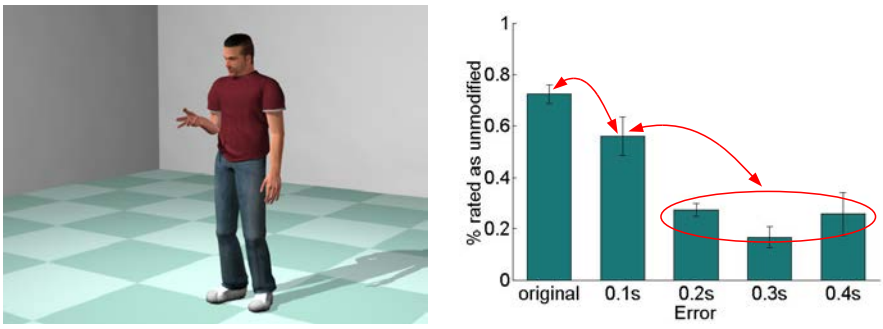
Data Analysis

A two factor (motion type and level of synchronization error) analysis of variance with repeated measures showed main effects of motion type ($F(3,60)=15.4$, $p<0.001$) and of synchronization error ($F(4,80)=17.6$, $p<0.001$). There was also an interaction effect between motion type and synchronization error ($F(12,240)=6.6$, $p<0.001$). We then performed a post-hoc analysis using standard Newman-Keuls tests for pairwise comparisons among means for each motion type.

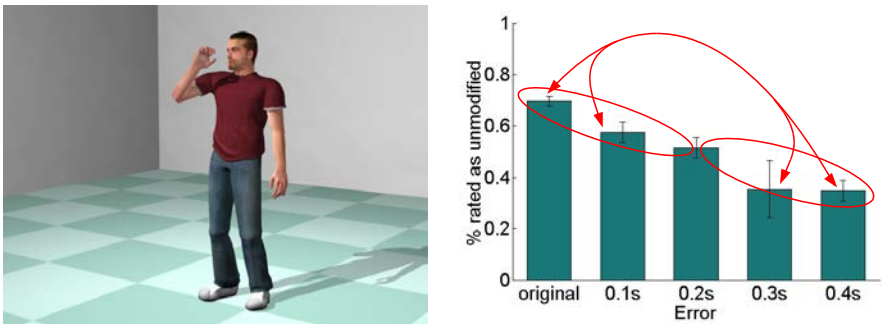
As can be seen from the results summarized in Figure 5.2, for the Count motion, there were significant differences between the original motion and any modified clip, including the motion with an error of 0.1 seconds. Furthermore, there were significant differences between the versions desynchronized by 0.1 seconds and all versions with greater error. In general, the error was detected more often for increasing error levels. A similar result, albeit to a lesser degree, was found for the Drink motion, where the synchronization errors could be recognized for errors bigger than 0.3 seconds. However, for the Snap motion a delay of 0.4 seconds was not detected at a significant rate, whereas a smaller error of 0.1 seconds was. Finally, for the Point gesture there was no significant effect of desynchronization level at all.

Discussion

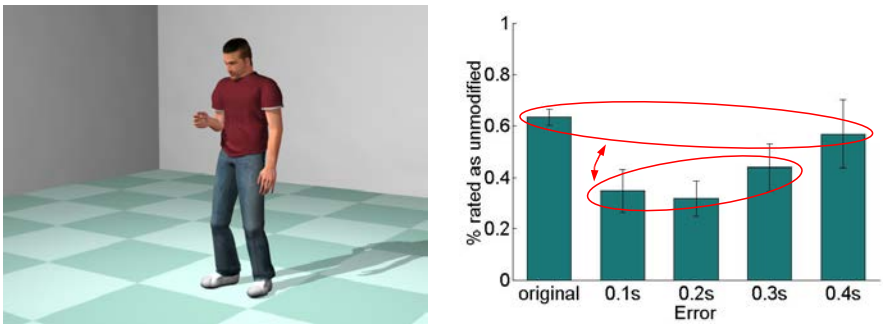
A reason for the observed results for Snap and Count could be the cyclic nature of those motions. Count has a longer cycle than Snap, which explains why the percentage of clips rated as unmodified does not increase significantly within the analysed error levels. The recognition rate was the highest for Snap, which had the highest velocity and energy of the fingers. This observation suggests that velocity or energy violations might be a factor in the perception of errors in motions. It would be very difficult to carry out a snapping action with desynchronized finger movements, whereas it is physically easy to perform a pointing action with delayed finger motion. This fact might explain why there was no significant effect of the synchronization errors at all for the Point gesture: all generated pointing motions were technically possible. A second reason why no errors were detected for the Point gesture might be that it was the shortest of all clips, leaving not enough time to detect the modifications. Errors in the Drink gesture were recognized at a significant rate when the error was greater than or equal to 0.2 seconds. As the motion was not as



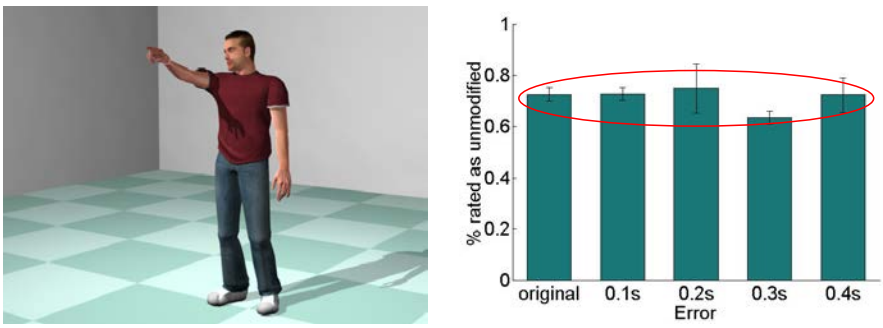
Count



Drink



Snap



Point

Figure 5.2: Screenshots and results for the gestures Count, Drink, Snap, and Point. The graphs show the percentage of motions rated as unmodified for the original motion and an increasing level of error. Values that are not significantly different from each other are circled, whereas significant differences are indicated with arrows. Error bars represent one standard error of the mean.

quick and energetic as the Count and the Snap gesture, but still more so than the Point motion, this supports our assumption that velocity or energy violations might influence the detection. In summary, we can say that there are significant differences in the ratings for errors as small as 0.1 seconds, but that this number depends on the gesture.

5.1.2 Comparing Finger Animation Methods

In a production environment, there is often a trade-off between desired animation quality and time and budget constraints. To provide some guidance, we evaluate the quality of four finger animation techniques: motion capture, keyframe animation, small random movements, and no motions at all. With a pairwise comparison task, we find that the motion captured clips are perceived to have the best quality.

Stimuli

A male actor was recruited and asked to narrate a common topic, in our case the content of a television series. His body and finger motions were captured simultaneously, using 20 markers attached on each hand and 48 on the body, with a 13 camera Vicon optical system (see Chapter 3.1 for details on the capturing process). From a 2 minutes 20 seconds motion captured narration, we chose 4 sequences of 3–6 seconds duration each, mostly consisting of beats and cohesive gestures that accompany speech. Finger motions in these types of gestures are quite subtle. They do not exhibit a specific meaning as the motions chosen in our first experiment do, but are used to punctuate and structure the speech.

For each sequence we created four types of finger animations. In the *original* condition the motion captured finger movements were displayed unaltered. Motion captured movements are without a doubt the most complex and resource intensive method that we tested. In contrast, the simplest one is the *no hands* condition, where the fingers were kept immobile in a natural resting position throughout the complete clip. To create the *keyframe* condition, we animated each sequence using up to six keyframes per hand. A skilled animator can create completely realistic hand motions with keyframes, the only limiting factor being time. We wanted to compare methods used in a typical production process with limited resources. Basic animations with only a few poses can be generated reasonably quickly. Therefore, we limited the number of keyframes to six per clip for each hand and the number of poses to three: one with the fingers in a relaxed position, one where only the index

was extended while the remaining fingers were kept relaxed, and one with all fingers extended. We then animated each sequence, while aiming to replicate the motion captured movements as accurately as possible within these constraints. The motion between the keyframes was interpolated using flat tangents. Lastly, the *random* condition examined if small movements, even when incorrect, are preferred over the unnatural case of completely stiff fingers. We cut the keyframe curves of the finger joints of our full animation in two parts and interchanged the two halves. As the original scene contains mainly subtle finger motions, we obtained small random motions of the fingers.

Each animation was displayed and rendered with the same character and environment as in our first experiment. In addition to the long shots with the same camera position as in Figure 5.2, we also rendered close-ups, in which the character was visible from the hips to the neck (see Figure 5.3). Because the face of the character is not visible in the close-ups and the hands can be seen more closely, we assumed that the differences between the conditions would be perceived to a greater extent. We obtained 4 (sequences) x 2 (shot types) x 4 (conditions) = 32 different animations, which we showed in pairs, keeping the sequence and shot type constant.

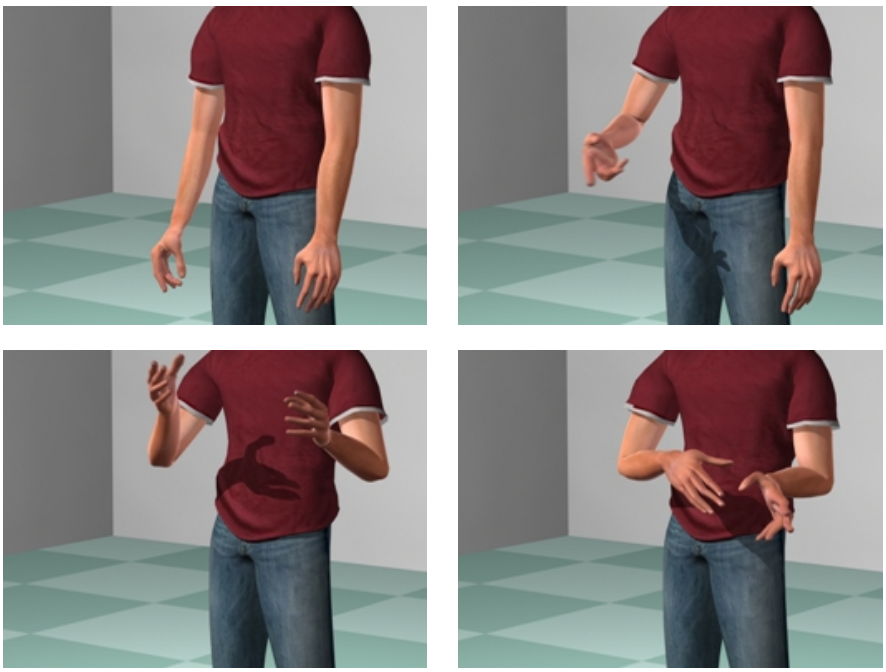


Figure 5.3: Frames from conversational gesture used to compare the quality of animation methods (motion captured movements, close-up view). From top left to bottom right: frames 1, 34, 84, and 103.

Participants

Twenty-four participants (10f, 14m, aged 21-45) viewed the clips in 6 groups. The participants were mostly students from a variety of disciplines and were recruited through e-mail lists and advertisements in college. Each participant was compensated with a 5 euro book voucher.

Procedure

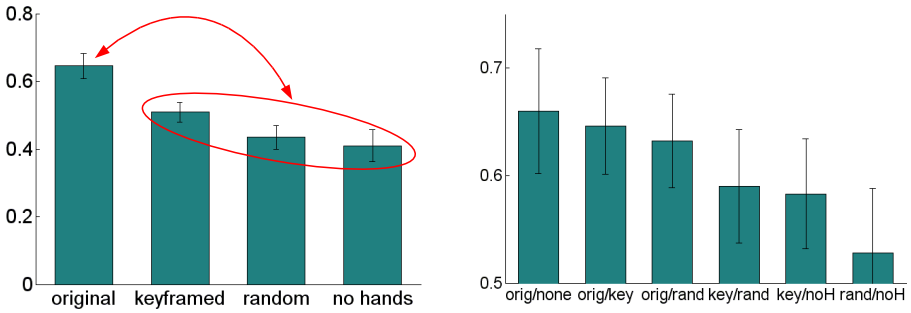
We used a two-alternative forced choice design where we compared all conditions against each other. So the participants were shown 6 (condition pairs) x 4 (sequences) x 2 (shot types) = 48 pairs of clips. Both, the order of the 48 pairs and the order of the two clips within a pair, were randomized for each group. For each clip we asked: *In which clip was the motion of better quality?* Similar to the procedure in the first experiment, a black screen was displayed for 5 seconds between each animation, during which participants noted their answer. After 3 seconds, they were alerted to the next stimulus by a sound. The experiment took about 20 minutes in total.

Data Analysis

During debriefing it became clear that many participants misinterpreted one of the sequences, in which the performer moved his middle finger independently of his other fingers. Many participants reported this particular motion as unnatural, so we removed it from the evaluation. We merged the results of the three remaining sequences and two shot types and carried out statistical analysis in order to verify the significance of overall preferences. The results are summarized in Figure 5.4. As expected, the original motion was preferred overall. In total, participants rated the original clip 64.6% of the time as being of better quality. The keyframed, random, and no hands conditions were selected 50.9%, 43.5%, and 41.0% of the time, respectively. T-tests for comparing the means of dependent samples showed that the original condition was preferred significantly more often than each of the three other conditions (all $p<0.05$), with no significant differences between the keyframed, random, and no hands animation. The difference between the keyframe animation and the no hands condition failed to reach significance.

The analysis of the pairwise preferences with t-tests for comparing the means provided

consistent results, as can be seen in Figure 5.4 (b), with the preferences for the three pairs that include the original condition being significant.



(a) (b)

Figure 5.4: (a) Overall preferences of animation method. The original animation is preferred significantly over the three other methods, whereas there is no significant difference between keyframed, random, and no animation (indicated with the circle and the arrow). (b) Pairwise preferences of animation method. The choices for the three pairs that include the original condition are significant. In both graphs error bars represent one standard error of the mean.

Contrary to our expectations, the viewpoint, i.e. long shot over close-up, had no significant effect on the perceived quality of the clips.

Discussion

Of the four conditions tested, the original motion-captured movements were judged as being qualitatively the best, which is what we expected. However, although we had predicted further preference differences between the three other conditions, these were not significant. A closer look at the individual participants showed that, out of those who had strong preferences for one condition over another, there was not always consensus amongst the participants on which condition had the better quality. For example, two participants judged the no hands condition to have the best quality consistently throughout the whole experiment (selected at least 5 times out of 6 for each pairwise comparison). Furthermore, as explained above, we found that even captured motions can look unnatural if the performed motion is not as expected. In such cases, the idealized motions of keyframed animations might be preferred.

It is noteworthy that there was no significant difference between the close-ups and the long shots, suggesting that finger motions may be equally salient in near and far shots (up to a certain limit). Apparently viewers could see the hand motions well enough in both viewpoints. The participants were not distracted by the face of the character, probably

because this was an experimental setting and they were told to focus on the hands.

5.2 Dimensionality of Finger Motion

As shown in the previous section, even subtle errors in hand motion can be perceived and it is challenging to generate qualitatively good hand motions with restricted time and budget. In this section, we present our method to explore the dimensionality of finger motions in order to reduce the number of degrees of freedom that need to be captured. We use root mean square (rms) deviations as a metric to determine which joints are moved in a similar way, so that we can express one as a linear transformation of the other. A scatter plot, shown in Figure 5.5, of the flexion of two joints indicates the existence of a linear relationship between two of the rotation coordinates, which implies that the motion of one of the joints can be approximated by a linear transformation of the other joint's motion. Depending on the goal, it is possible to choose the requested level of dimensionality reduction. We implement different levels of simplification, and we discuss the resulting motions.

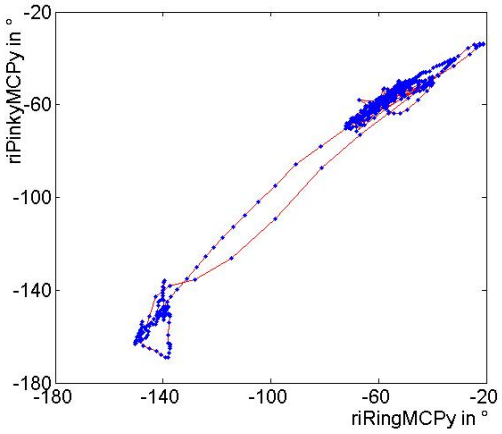


Figure 5.5: Scatter plot of the flexions of the right ring finger's MCP-joint and the right little finger's MCP-joint indicates a linear relationship between the joints' motions. Each point represents a frame; successive frames are connected by a line.

Hands are used for manipulative tasks where the digits interact with an object or a person, as well as for gestures that might or might not include interaction or self-interaction. Hand motions that include contacts depend to a large degree on the task and on the object of interaction. They cannot be generated accurately without knowledge of the task. We therefore exclude any type of contacts and focus on conversational gestures. For this

exploratory analysis we use the hand motions of a male actor during 27.3 seconds depicting a staged argument.

The skeleton model presented in Figure 3.3 on page 34 is used. We denominate a joint on the right hand with *ri* and a joint on the left hand with *le*. The little finger is denoted by *Pinky*. Thus, *riRingDIPy* characterises the curve of the right ring finger's DIP-joint's y-rotation, i.e. its flexion or extension. *lePinkyMCPx* denotes the curve of the left little finger's MCP-joint's x-rotation, i.e. its abduction or adduction. Each rotation coordinate is represented as a sequence of frames with 30 frames per second. The rotations are specified in degrees, relative to their parents in the hierarchy. The animation consists of 27.3 seconds or 820 frames with a keyframe at every frame. In total, taking the constrained degrees of freedom of the joints into account, this sums to 25 rotation curves for each hand.

5.2.1 Proposed Method

Our proposed method is to use Root Mean Square (RMS) deviations as a distance metric in order to identify which joint rotations are most strongly correlated. We then use the rotation curve with the larger standard deviation to approximate the curve with the smaller standard deviation, using a linear model to summarize the relationship between the two curves. We also remove rotations with a small range as they are barely visible. In summary, our technique takes advantage of redundancy (correlations) and irrelevance (barely visible rotations).

Our algorithm consists of the following rules:

1. If the range of a joint rotation is smaller than a specified threshold, we replace the angular values of each frame with the mean value of the whole motion.

2. If the root mean square deviation of two joint rotations is less than a threshold, the curve a with the smaller standard deviation is created based on the curve b with the larger standard deviation in the following way: $a = (b - mean(b))/std(b) * std(a) + mean(a)$.

3. If the root mean square deviation of one joint rotation and of the inverse of a second joint rotation is less than a threshold, the curve a with the smaller standard deviation is created based on the curve b with the larger standard deviation in the following way: $a = (mean(b) - b)/std(b) * std(a) + mean(a)$.

The thresholds can be specified depending on the application. We present examples of different thresholds in Section 5.2.2. Once the relationships between curves have been established, new hand motions can be created while only the remaining degrees of freedom are recorded.

Find irrelevance

First, we calculate the range, the mean, and the standard deviation for each curve. Figure 5.6 shows the distribution of the ranges of the rotation curves. As can be seen in the histogram, the majority of the ranges are small. Out of 50 joint rotations, 19 have a range of less than 5°; for 11 coordinates the range is even less than 1°. As a rotation of 1° is barely visible, we can delete those rotations without losing too much information. We replace the curves of the 11 joints that have a range of less than 1 with a constant value: the mean of the curve. A rotation of 5° is visible, but it does not change the meaning of the gesture. In a later step, we replace the coordinates that have a range of less than 5 in order to reduce the degrees of freedom of the hands even further.

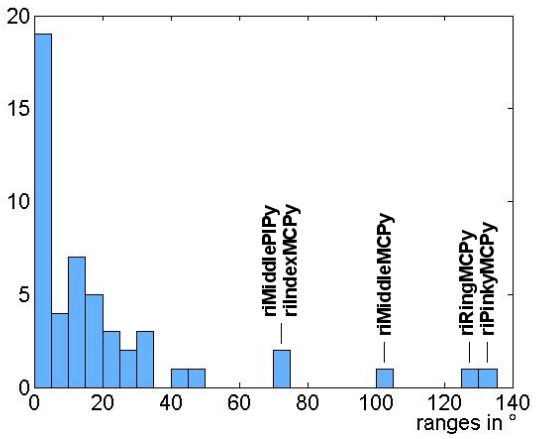


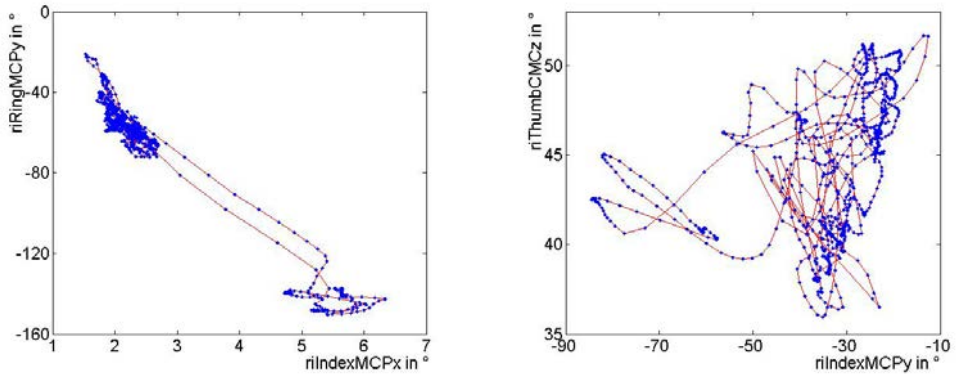
Figure 5.6: Distribution of the ranges of the 50 rotation curves representing the hand motions.

Find correlations

Figures 5.5 and 5.7 show three exemplary scatter plots in each of which two curves are plotted against each other. We can see that a line with either a positive or negative slope is a good approximation for the relationship between some pairs of rotation curves.

To find which curves are most strongly correlated, we use the RMS deviation as a distance

metric. First, we normalize each rotation curve by subtracting the mean from the value at every frame and dividing by the standard deviation, resulting in a set of curves that all have mean 0 and standard deviation 1.



(a) (b)

Figure 5.7: Scatter plots of two pairs of rotation curves. Each point represents a frame; successive frames are connected by a line. (a) The relationship between riIndexMCPx and riRingMCPy can be approximated by a line with negative slope. (b) A line would not be an accurate approximation for the relationship of riIndexMCPy and riThumbCMCz.

To detect values that are highly positively correlated, we calculate the square deviations of the corresponding points of two curves and compute the root of the mean of those deviations. Let $a = (a_1, a_2, \ldots, a_n)$ and $b = (b_1, b_2, \ldots, b_n)$ denote two curves with n the number of frames. Then our distance metric is calculated as:

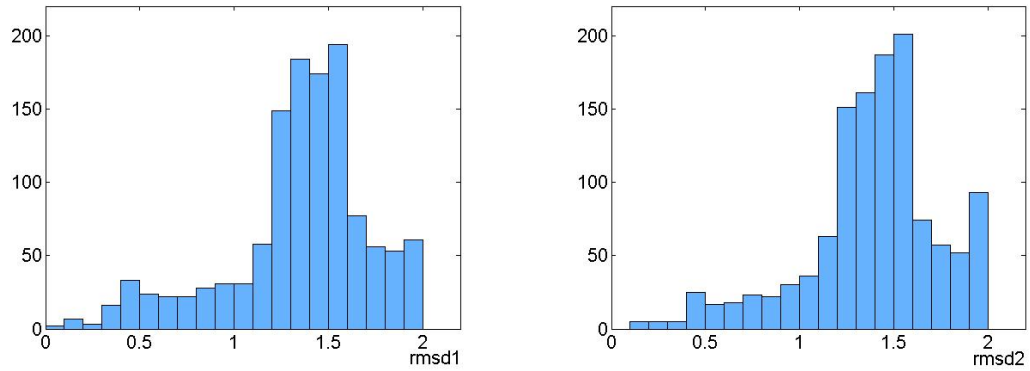
$$rmsd_1(a, b) = \sqrt{\frac{\sum_{i=1}^{n}(a_i - b_i)^2}{n}}$$

We repeat this process for every pair of curves, resulting in $50 * 49/2 = 1225$ values. Figure 5.8 (a) shows the distribution of those values. The smaller the value of this distance metric, the more accurately one curve can be estimated as a linear transformation of the other curve, the slope of the line approximating the relationship in a scatter plot being positive.

For values that are highly negatively correlated, we adapt the distance metric to calculate the RMS deviation of one curve with the second curve mirrored across the x-axis.

$$rmsd_2(a, b) = \sqrt{\frac{\sum_{i=1}^{n}(a_i + b_i)^2}{n}}$$

Again, we compute this metric for every set of two rotations. Figure 5.8 (b) graphs the



(a) (b)

Figure 5.8: (a) Distribution of the 1225 values for rmsd_1, the distance metric for linear relations approximated by a line with a positive slope. (b) Distribution of the 1225 values for rmsd_2, the distance metric for linear relations with a negative slope.

distribution of the 1225 resulting values. Analogous to rmsd_1, the smaller the value of this distance metric, the better one curve can be estimated based on the other one. In this case the slope of the line is negative.

Simplify curve using rmsds

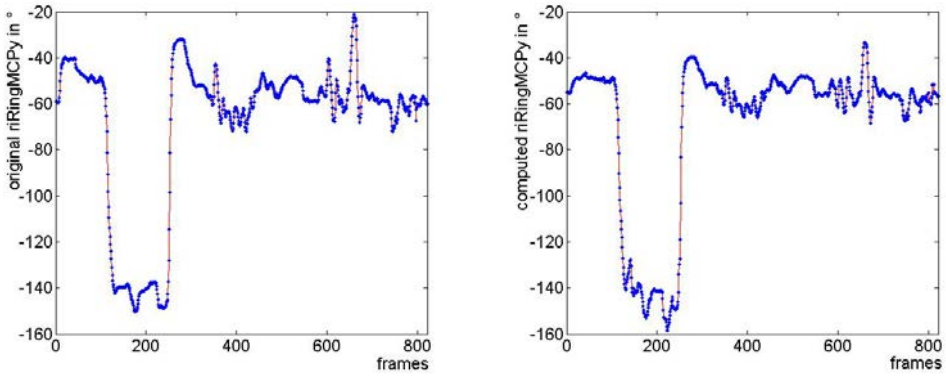
When the distance metric is less than a chosen threshold, we compute the curve with the smaller standard deviation as a linear combination of the curve with the larger standard deviation using the following equations for each point i, with $1 \leq i \leq 820$, for rmsd_1 and rmsd_2, respectively.

$$a_i = \frac{b_i - mean(b)}{std(b)} * std(a) + mean(a)$$

$$a_i = \frac{mean(b) - b_i}{std(b)} * std(a) + mean(a)$$

This computation results in curves having the same mean and the same standard deviation as the original curve.

Figures 5.9 (a) and 5.9 (b) show an original rotation curve and the same rotation computed as a linear combination of a different curve.



(a) (b)

Figure 5.9: (a) Original rotation curve riRingMCPy. (b) Rotation riRingMCPy computed as a linear combination of the curve riPinkyMCPy.

5.2.2 Results

In this study we present four steps of simplification based on the following criteria:

1. range < 1
2. $\text{rmsd}_1 < 0.3$ or $\text{rmsd}_2 < 0.3$
3. range < 5
4. $\text{rmsd}_1 < 0.5$ or $\text{rmsd}_2 < 0.5$

Step 1 and 3 affect rotation curves with small ranges, step 2 and 4 are based on the proposed distance metric.

Step 1: range<1

The 11 rotation curves that have a range of less than 1, as well as their ranges and means, are listed in Table D.1 in the appendix.

Each curve is replaced by a constant value: its mean. The listed rotation curves either describe the flexion (y-axis) of the DIP-joints or represent adduction, abduction (x-axis) or rotation motions (z-axis) of the left hand. These joint rotations are expected to be small. The left hand is moving less than the right hand in the analysed animation sequence, which explains why the majority of the simplifications affect the left hand.

Step 2: $\mathbf{rmsd}_1 < 0.3$ or $\mathbf{rmsd}_2 < 0.3$

Once the curves with a range less than 1 are excluded, 11 rmsd_1 values and 10 rmsd_2 values are less than 0.3. However, within these relationships, there are overlaps, e.g. if curve a is controlled by curve b, but curve b is already controlled by curve c. In this case we compute curve a based on curve c, even if the distance metric exceeds the threshold (this happens for only two joint rotations in Step 4). If the distance metric of curve a and curve b as well as the distance metric between curve a and curve c are less than 0.3, curve a could be controlled by curve b or curve c. We then calculate curve a based on the curve with the larger standard deviation. Table D.3 in the appendix lists the remaining relationships after simplification according to the above criteria.

These 12 simplifications – one is represented in Figure 5.9 – cover a diverse set of joint rotations. We are therefore able to retain variation within the motion.

Step 3: range<5

Eight rotation curves have a range between 1 and 5 and are therefore simplified in this step, as shown in Table D.2.

Similar to Step 1, these values continue to reduce the flexion (y-axis) of the DIP-joints and adduction, abduction (x-axis) and rotation values (z-axis), which all have low ranges in natural human hand motion. An exception is the PIP-flexion of the right little finger. The small range of this joint rotation is probably due to approximation within the skeleton model, which does not take the motion of the fifth carpometacarpal joint allowing the bending of the palm on the side of the little finger, into account. This simplification causes an increase in the range of the MCP-joint of the little finger and a decrease of the range of the PIP-joint.

Step 4: $\mathbf{rmsd}_1 < 0.5$ or $\mathbf{rmsd}_2 < 0.5$

In a fourth step, we choose all pairs with $\text{rmsd}_1 < 0.5$ or $\text{rmsd}_2 < 0.5$. Table D.4, which lists the added or changed relationships, shows that, after the computation, most joint rotations only depend on a few original rotations: PinkyMCPy of both hands and rotations of the left middle finger and both thumbs.

Summary

Within four different stages, we continually reduce the degrees of freedom of the hand. The amount of remaining DOFs are listed in Table 5.1.

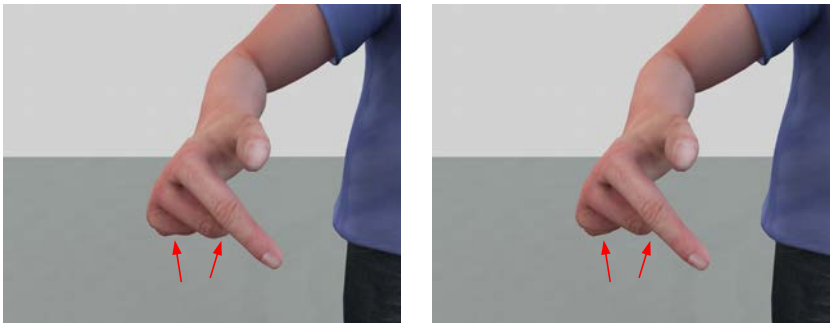
condition	DOFs left
original	50
range< 1	39
range< 1 + rmsd< 0.3	27
range< 5	31
range< 5 + rmsd< 0.5	15

Table 5.1: Remaining degrees of freedom after each step of dimensionality reduction.

From the original 50 DOFs for both hands we reduced the motions to 15 DOFs, which shows the potential of our approach. We applied our results and rendered our original animation with only 15 DOFs.

5.2.3 Discussion

Exemplary results can be seen in Figures 5.10 and 5.11. A comparison of the figures shows only marginal differences between the original version and the animation taking only 15 DOFs into account.

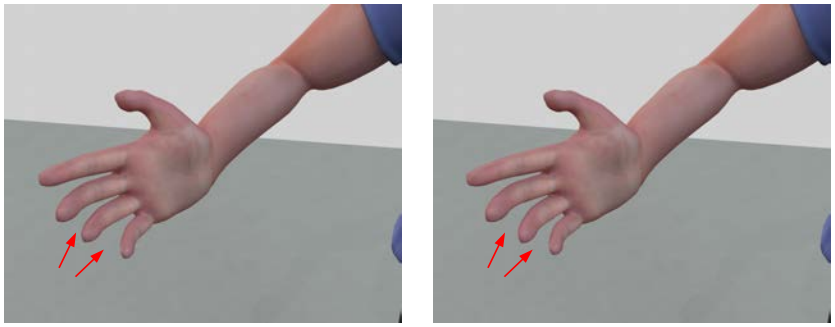


(a) (b)

Figure 5.10: Close-up of the hand of a frame from the animation: (a) original version; (b) most simplified version (after step 4). Small differences can be seen, for example, when looking at the little finger.

Our results show that there are no strong correlations between the motions of the right hand and the motions of the left hand. In both hands, many adduction/abduction or rotation curves of the DIP-joints are reduced. This is not surprising as the range of those rotations is small in natural human hand motion.

For the right hand, numerous joints' flexions of the middle, ring and little finger are created



(a) (b)

Figure 5.11: Close-up of the hand of a frame from the animation: (a) original version; (b) most simplified version (after step 4). Small differences can be seen, for example, when comparing the spaces between the fingers.

based on the flexion of the MCP-joint of the little finger, which has the largest standard deviation. This seems plausible: Firstly, the range of motion of the flexion of the MCP-joints increases from the index to the little finger in a human hand. Secondly, the chosen hand model is likely to further increase this difference. The CMC-joint of the thumb is completely independent from the other digits and keeps its three degrees of freedom. Further curves that are independent are two rotation curves of the index as well as the *riMiddleMCPz* rotation. A closer look at the latter shows that the range of this joint is 7.1 only because of a few outliers in fewer than 10 frames. Without these outliers, the range would be less than 4.0 and the curve would have been simplified in the third step.

The correlations are less explicit for the left hand, which keeps more variation in the motion. The flexions of the fingers are partly based on the middle finger's PIP- and MCP-joints, and partly on the little finger's MCP-joint. Similar to the right hand, the motion of the thumb is independent, keeping three DOFs. Furthermore, the rotations *leIndexMCPy* and *lePinkyPIPy* are kept unchanged without any curve being dependent on them.

The rotation curves that become simplified are similar for the left and the right hand, but not exactly the same. We therefore assume that our findings may vary slightly but not substantially depending on the handedness of the actor.

An alternative approach to reduce the dimensions would be to compute correlation coefficients of pairs of curves as a distance metric, which leads to very similar results. More sophisticated techniques such as PCA may provide interesting results, though they are likely to yield an unintuitive set of basis poses. Nevertheless, more complex models that retain basic rotations could include relationships with more than one joint rotation and (self-)relationships within time.

If oversimplified, for example when nearly all of the rotation curves are generated based on just one curve, the motion could seem artificial. Using a larger variety of relationships would add more variation to the resulting motions. Furthermore, it is possible to add plausible variation into the rotations by adding noise.

The initial model of the hand, which influences the calculations of the rotation angles, is a simplification of the complex structure of a real hand. Even though the chosen simplifications are common in character animation, they alter the real motions of the hand. Simplifications compared to a real human hand occur, especially at the palm, and influence the motions of the little finger and the thumb. A more accurate model would allow for a more exact analysis.

In summary, we advise an animator or a person capturing motions to concentrate on the thumb and on the MCP-flexions of the index and one additional finger, such as the middle or pinky, as their rotation curves convey the most essential information of the motion.

5.3 Adding Finger Animation to the Metropolis System

Metropolis is a real-time system that simulates Trinity College Dublin and its surroundings with its streets, crowds, traffic, and sounds. Furthermore, Metropolis provides a testbed for perceptual experiments. We establish a pipeline to integrate finger motions into the Metropolis system. No new finger motions were generated for the system as part of this dissertation. The pipeline is provided only to enable the future addition of finger motions.

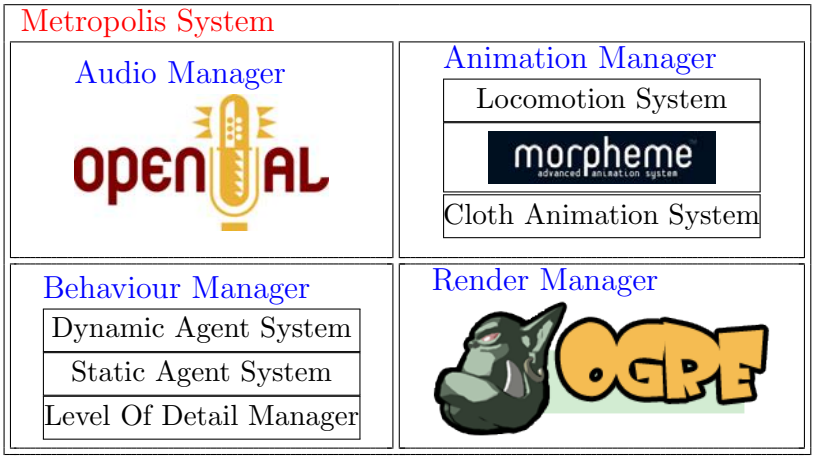


Figure 5.12: Metropolis System overview.

The system consists of an audio, behaviour, animation, and render manager (see Fig-

ure 5.12). The free audio API OpenAL (Open Audio Library) developed by Creative Technology [Ope11] is used as the audio manager as it is able to render three dimensional positional audio. The behaviour manager consists of a dynamic agent system carrying out the path planning and steering behaviour, a static agent system, which controls social groups, and the level of detail (LOD) manager, which handles the choice of the appropriate level of detail depending on the distance from the camera and the visibility. A proprietary locomotion system, that is able to handle several thousand pedestrians, the morpheme animation system from Natural motion [Nat11] that provides hierarchical finite state machines (FSM), and a cloth animation subsystem form the animation manager. The 3D rendering engine OGRE (Object-Oriented Graphics Rendering Engine) [OGR11] is used for real-time rendering.

Our goal is to add finger animations to the Metropolis System. In the Animation Manager, the Locomotion System drives the walking characters and the FSM system Morpheme controls groups of standing, conversing characters (see Figure 5.13). When walking, our hands mostly move by small amounts only. Furthermore the arms are swinging, which makes the details in finger motions hard to detect. Hand animations for walking characters would probably only increase the quality of our system marginally while spending valuable computing time. We hence concentrate our efforts on the animation of the fingers for conversing characters. This means that we need to add finger animations to the characters in the Morpheme System.

Figure 5.13: Screenshot from the Metropolis System with a conversing group in the foreground and walking agents in the background.

Contrary to the pipeline for finger motions presented in Section 3.1 where we used Vicon to calculate skeleton orientations and Maya to render animations, the character motions for Metropolis are generated with Vicon and 3ds Max. The software 3ds Max developed by Autodesk [Aut11] is a modelling, animation, and rendering package. For character animation it features a pre-built skeleton, called “biped”. The motion captured motions

are only labelled in Vicon and exported as a CSM animation file. In 3ds Max, they are loaded onto a biped, from where they can be retargeted automatically to different models. The characters are then repositioned and the bones renamed to adapt all characters to the Morpheme naming conventions. Finally, character meshes and animations are exported and integrated into a finite state machine in Morpheme. The FSMs describe all the different possibilities of animations that can be played one after another.

To integrate finger animations into this pipeline, we used scripting languages of Maya and 3ds Max. Maya supports Python scripting while 3ds Max provides the built-in scripting language MAXScript. We developed a script in Python that extracts the Euler rotation coordinates for each finger joint and each frame from an animation in Maya and saves them as a simple text file. In the same way, a script was generated in MAXScript to save each rotation value as a keyframe. Joint names and rotation order had to be adapted to fit each software accordingly.

Now that this pipeline is established, it is possible to generate different types of finger motions and to add variation to conversing characters in the Metropolis System in future work. We furthermore open up the study of further questions, for example what the perceptual consequences of different types of finger motions are on conversing groups of characters.

Chapter 6

Conclusions and Future Work

Realistic human-like characters are challenging to create and the first completely realistic virtual character still needs to be brought to life. While motion capture provides the possibility to generate very natural motions, flaws still exist in this technique. Human observers are adept at interpreting subtle differences in motions and linking this information with stimuli from other senses, such as sound. The underlying perceptual processes are still not well understood and current research methods are limited in their ability to test human perception.

Observing human reactions to controlled stimuli in an experimental setting is an effective method of enhancing our understanding of these perceptual processes. This dissertation has presented a method to systematically investigate the effect of errors and degradations in character animation on the emotional response and interpretation of a scene. We focused on the impact of errors in animation that might happen when animating very realistic virtual humans using motion capture. Based on our results, we provided experimental evidence and advice on how to improve a specific crucial aspect of human animation: the hands.

The primary contributions of this thesis can be summarized as follows:

First, we introduced methods to improve the accuracy of current optical motion capture systems to generate realistic hand animations with the available hardware. We provided advice on how to improve and optimize marker sets, camera configurations, the capture session and post-processing to capture precise motions.

Second, we presented a novel approach to evaluating the impact of errors and degradations

of realistic human-like characters on the viewer. This is the first study that analyses the reactions to errors in animations with longer stimuli (the vignettes) that are produced to be as realistic as possible. Using a survey questionnaire, we evaluated the impact of our stimuli on the emotional response, the attention to details, the interpretation, and the perceived quality of the animation by the viewer. We furthermore studied differences in gaze patterns. We successfully showed that changes in stimuli indeed lead to alterations in the perception of viewers. In addition, we also demonstrated the importance of randomization and a careful choice of participants in the experimental design.

The most important results of our perceptual experiments can be summarized as follows:

- It is possible to alter the interpretation of a scenario by introducing very subtle errors, even without changing the perceived quality of an animation (Computer Crash vignette with finger motions desynchronized by 0.5 seconds).

- Even very subtle errors can change the emotional response and the interpretation of a scenario. Furthermore contacts and interactions might be crucial in our processing of human motions (Argument vignette with the movements of one character delayed compared to the other character by 0.17 seconds).

- Degradations in character animation can alter the perceived quality of an animation. The relationship between perceived quality and human likeness is not linear. More precisely, a U-shaped relationship similar to that of the uncanny valley was found, meaning that a more stylized character might be preferred over a more human-like one.

- People gaze significantly more often at the faces of virtual characters compared to their bodies. As a result, facial anomalies are particularly salient, even when considerable body anomalies are present.

- The interactions of sound and degradations in animation on the perception of human-like characters is complex and not fully understood.

Next, in two experiments, we provided insights into the perception of anomalous finger movements on virtual humans. We showed that very subtle errors can still be detected, such as a 0.1s desynchronization error. However, we also provided strong evidence that the perceptibility of this type of error is significantly affected by the type of action being performed, as for one test motion even errors of 0.4 seconds went undetected. Furthermore,

we demonstrated the difficulty of producing finger animation in an automatic way, as none of our three tested animations – no motion at all, keyframed animation and randomly selected motion captured movements – reached the perceived quality of synchronously captured movements.

Furthermore, we analysed conversational hand motions to identify barely noticeable movements and rotations that can be accurately computed as a linear relation of each other. Based on our findings, we reduced the dimensions of the motions of digits while keeping an intuitive base of rotations. The transformations are simple and do not require any elaborate computations. We showed that we can describe the finger movements with 15 degrees of freedom in total for both hands with barely noticeable differences. Based on our results, we advised an animator to concentrate on the thumb and on the MCP flexions of the index and one additional finger, such as the middle or pinky, as their rotation curves convey the most essential information of the motion.

Finally, we presented a pipeline to integrate hand motions into a crowd animation system, specifically to groups of conversing characters in that system.

In summary, this thesis contributes to both substantive as well as methodological questions related to the evaluation of human-like characters, paying particular attention to improving hand animations as one crucial aspect of the animation of virtual humans. Future research efforts could be guided by the advice and findings provided in this study when developing new experiments to analyse the perception of human motions, and when trying to create realistic virtual characters. The findings presented in this thesis, however, go beyond the areas of perception and computer graphics.

Our contributions on the effect of errors in realistic human-like characters might be applied to the study of human interaction in psychology. They could help to support the social development of autistic children similar to current experiments with robots [KNY05, FsM08], where less realistic robots are used to bond with the child. The realism of virtual characters could be varied to different levels step by step until reaching a realistic human-like character. Furthermore, our results could be applied to virtual reality applications and in educational or training settings. Our findings on the dimensionality of finger motions might be useful in the field of robotics to create robots able to perform flexible grasps. The knowledge of correlations in finger motions could be useful for computer vision applications. The pipeline to animate the fingers of virtual characters can be used to generate

new experiments and to gain further insight into the perception of fingers in conversing characters.

The findings in this thesis point to new and exciting directions for future research. It would be interesting to test whether the small set of vignettes used in our experiments can be applied to other scenarios. In addition, the reasons why we found effects and why we did not in some vignettes have to be further investigated. Additionally, further study is needed to understand the interactions between audio and visual stimuli. As it is common in practice to capture motion and audio tracks separately (e.g. for performance capture with a VIP or A-List actor, or when animated movies are dubbed in another language), if the emotional content is being disrupted by the bodies being more or less emotional than the voices or desynchronized due to bad dubbing, this is worthy of further investigation. In a recent study Carter et al. show that incorrect audio synchronization does affect the perceived emotional intensity of a performance [CST*10]. Now that we have established a framework for exploring these factors with longer, more natural stimuli, we can consider developing more detailed or sophisticated questionnaires to explore them further across a broader range of stimuli. Our framework can now be extended to investigate, for example, differences in responses to cartoon rendering, video, and realistic rendering. With longer vignettes, we could consider using "sliders" that allow people to record the perceived emotions over time, and not just at the end of a vignette. Physiological measures such as heart-rate, brain responses and more in-depth eye-movement analysis could also provide insight. Finally, an interesting question is whether the same effects are found when anomalies occur in less familiar, non-human characters, as in *Avatar*.

In the perception of finger motions, it would be interesting to determine which features of a motion contribute to the perceptibility of synchronization errors. Do velocity, contacts, energy violations, or familiarity play a role? Are those results transferable to other virtual characters, e.g., more or less realistic ones? Hodgins et al.'s study [HOT98] suggests that this might not be the case and it would be interesting to replicate their study for finger motions. Do further factors, such as the quality of the rendering, the lighting, or the complexity of the scene, influence the results? Speech and gestures are highly correlated, so would the presence of audio and lip synchronization alter our results? Finally, how can we create finger motions of good quality in a more automated way than with motion capture? To answer this question, we would like to validate the proposed dimensionality reduction approach with a wider range of motions and actors. Further work is clearly

needed to understand the complexity of human perception of finger animation. Conversational agents or the pipeline that we described to integrate finger movements into a crowd animation system would be valuable for that purpose. A better understanding of finger movements and their perception would be useful to generate gestures and finger motions fully automatically based on the body motion or on text input for conversational characters. A promising way forward would be to develop new techniques that further simplify and automatize the animation of fingers.

Appendix A

Gesture Sequence for Hand Range of Motion

The following list of hand poses was used to generate an accurate calibration of the hand vsks. The fingers were bent by 90° or as far as possible. A book was used to help bend the joints individually.

1. Put all fingers straight with the dorsal surface of the hand facing the ceiling → T-Pose
2. Turn wrist to make thumb point upwards
3. Bend all interphalangeal joints and the metacarpophalangeal joint of the thumb
4. Back to T-Pose
5. Bend thumb
    - carpometacarpal joint
    - metacarpophalangeal joint
    - interphalangeal joint
6. Bend metacarpophalangeal joints one by one
    - index
    - middle
    - ring

- pinky

7. Bend proximal interphalangeal joints one by one
   - index
   - middle
   - ring
   - pinky
8. Bend distal interphalangeal joints one by one
   - index
   - middle
   - ring
   - pinky
9. Back to T-Pose
10. Join the fingers except the thumb, bend the metacarpophalangeal joints of all fingers at the same time
11. Back to T-Pose

Appendix B

Synchronization Errors Questionnaire

The next pages contain the original questionnaire that was used in the experiment presented in Section 4.1, including the instructions that were handed out to participants. The goal of this experiment was to study the effect of synchronization errors on the perception of emotional content, attention to details, interpretation, and perceived quality. The questionnaire contains questions for each of the three vignettes that were used (Computer Crash, Bus Wait, Argument), followed by questions about the quality of the animation and general questions about the participants.

Instructions and Questionnaire

In this experiment you will view three short videos and will be asked to respond to a couple of questions about those videos.

This experiment is being conducted in connection with a research project on virtual humans. Your participation is fully voluntary, and you may decline to participate in the experiment at any time. Your data will not be used for any other purposes.

By signing below, you indicate that you agree to participate in the study.

Signature: _____

Computer

1. How angry would you rate the character in the animation?

not angry				very angry
☐	☐	☐	☐	☐
1	2	3	4	5

2. Is his reaction appropriate?

inappropriate				appropriate
☐	☐	☐	☐	☐
1	2	3	4	5

3. Does the scene occur during the day or during night time?

day	night
☐	☐

4. How often does the character try to move the mouse? _____

5. Please describe what happened in that scene. _____

_____

_____

_____

Waiting

1. How would you rate your sympathy towards the character in the animation?

low				elevated
☐	☐	☐	☐	☐
1	2	3	4	5

2. How nervous would you rate the character?

not nervous				nervous
☐	☐	☐	☐	☐
1	2	3	4	5

3. How impatient would you rate the character?

not impatient				impatient
☐	☐	☐	☐	☐
1	2	3	4	5

4. How many times does she check her watch? _____

5. What kind of building is in the background? _____

6. What colour are her trousers? _____

7. Why would you think is she so fidgety? _____

_____

8. Please describe what happened in that scene. _____

_____

_____

_____

Argument

1. Is the reaction of the man appropriate?

not appropriate				appropriate
☐	☐	☐	☐	☐
1	2	3	4	5

2. Is the reaction of the woman appropriate?

not appropriate				appropriate
☐	☐	☐	☐	☐
1	2	3	4	5

3. How angry is the man?

not angry				angry
☐	☐	☐	☐	☐
1	2	3	4	5

4. How angry is the woman?

not angry				angry
☐	☐	☐	☐	☐
1	2	3	4	5

5. Who pushes the other person?

woman	man
☐	☐

6. What colour is her jumper? _____ What colour is his jumper? _____

7. Who is responsible for that argument?

woman	man
☐	☐

8. Please describe what happened in that scene. _____

_____

_____

Quality of Animation

1. How would you rate the quality of the animation?

	low quality				high quality
Computer	☐	☐	☐	☐	☐
Argument	☐	☐	☐	☐	☐
Waiting	☐	☐	☐	☐	☐
	1	2	3	4	5

2. What made you score the animation as high or low quality?

   Computer _____

   Argument _____

   Waiting _____

General Questions

1. Are you:

male	female
☐	☐

2. Have you ever participated in a similar experiment?

yes	no
☐	☐

   If yes, what was that experiment about? _____

Thanks for your participation!

Please write any comments you have here: _____

_____

_____

Appendix C

Effect of Degradations Questionnaire

The next pages contain the original questionnaire that was used in the experiment presented in Section 4.2 on the effect of degradations on the viewer's perception of emotional content, attention to details, interpretation, and perceived quality. The questionnaire contains a preliminary page, questions for each of the three vignettes that were used (Milk, Money, Moving Out), followed by closing questions on all three vignettes.

Preliminary Questions

1. Are you:

male	female
☐	☐

2. How old are you?

under 21	21 - 25	26 - 30	31 - 40	41 - 50	51 - 60	over 60
☐	☐	☐	☐	☐	☐	☐

3. What is your field of work or studies? _____

4. What are your three favorite computer games? _____

_____

5. Did you watch the following movies and, if yes, how many points would you give each of them, 5 meaning you liked the movie a lot, and 1 meaning that you didn't like the movie at all?

	not seen	1	2	3	4	5
Beauty and the Beast (1991):	☐	☐	☐	☐	☐	☐
Pocahontas (1995):	☐	☐	☐	☐	☐	☐
Shrek (2001) :	☐	☐	☐	☐	☐	☐
The Polar Express (2004):	☐	☐	☐	☐	☐	☐
Wallace & Gromit (2005) in The Curse of the Were-Rabbit:	☐	☐	☐	☐	☐	☐
Ratatouille (2007) :	☐	☐	☐	☐	☐	☐
Beowulf (2007):	☐	☐	☐	☐	☐	☐
	not seen	1	2	3	4	5

6. What other movies with virtual or cartoon humans have you seen, if any? How many points would you give each of them, 5 meaning you liked the movie a lot, and 1 meaning that you didn't like the movie at all?

	1	2	3	4	5
_____	☐	☐	☐	☐	☐
_____	☐	☐	☐	☐	☐
	1	2	3	4	5

PLEASE DO NOT TURN PAGE UNTIL INSTRUCTED TO DO SO

Animation

1. How angry would you rate the characters in this animation?

	not angry						very angry
Woman:	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7
	not angry						very angry
Man:	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7

2. How appropriate would you rate the characters' behavior?

	very inappropriate						completely appropriate
Woman:	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7
	very inappropriate						completely appropriate
Man:	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7

3. How would you rate your sympathy towards the characters?

	very low						very high
Woman:	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7
	very low						very high
Man:	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7

4. How plausible would you rate the events shown?

very unrealistic						very realistic
☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7

5. Please describe what happened. _____

_____

_____

_____

6. Who is mainly responsible for the argument? woman ☐ man ☐

7. What were they arguing about? _____

8. Please describe their physical interaction.

_____

9. Which type of fruit is on the counter? _____

10. Is the coffee maker on the left or right side? left ☐ right ☐

11. What is the woman trying to do when the scene starts?

_____

12. What does the man do after the scene?

_____

PLEASE DO NOT TURN PAGE UNTIL INSTRUCTED TO DO SO

Animation

1. How angry would you rate the characters in this animation?

	not angry						very angry
Woman:	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7
	not angry						very angry
Man:	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7

2. How appropriate would you rate the characters' behavior?

	very inappropriate						completely appropriate
Woman:	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7
	very inappropriate						completely appropriate
Man:	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7

3. How would you rate your sympathy towards the characters?

	very low						very high
Woman:	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7
	very low						very high
Man:	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7

4. How plausible would you rate the events shown?

very unrealistic						very realistic
☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7

5. Please describe what happened. _____

_____

_____

_____

6. Who is mainly responsible for the argument? woman ☐ man ☐

7. What were they arguing about? _____

8. Please describe their physical interaction.

_____

9. What kind of picture is on the front of the fridge? _____

10. Was there a plant on the windowsill? yes ☐ no ☐

11. What did the man do just before this scene?

_____

12. What does the woman try to prevent the man from doing?

_____

PLEASE DO NOT TURN PAGE UNTIL INSTRUCTED TO DO SO

Animation

1. How angry would you rate the characters in this animation?

	not angry						very angry
Woman:	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7
	not angry						very angry
Man:	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7

2. How appropriate would you rate the characters' behavior?

	very inappropriate						completely appropriate
Woman:	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7
	very inappropriate						completely appropriate
Man:	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7

3. How would you rate your sympathy towards the characters?

	very low						very high
Woman:	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7
	very low						very high
Man:	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7

4. How plausible would you rate the events shown?

very unrealistic						very realistic
☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6	☐ 7

5. Please describe what happened. _____

_____

_____

_____

6. Who is mainly responsible for the argument? woman ☐ man ☐

7. What were they arguing about? _____

8. Please describe their physical interaction.

_____

9. What kind of beverage is on the counter? _____

10. Was there a vase on the shelf? yes ☐ no ☐

11. Why is the man there?

_____

12. What does the woman do next?

_____

PLEASE DO NOT TURN PAGE UNTIL INSTRUCTED TO DO SO

Closing Questions

1. How would you rate the quality of the animation from a technical point of view?

very low quality						very high quality
☐	☐	☐	☐	☐	☐	☐
1	2	3	4	5	6	7

2. What made you score the animation as high or low quality?

3. Putting the three movies in what you believe to be the right chronological order, please tell the full story surrounding the events depicted.

Thanks for your participation!

Please write any comments you have here:

Appendix D

Detailed Results for Dimensionality Reduction

Tables D.1 to D.4 detail the rotation curves with small ranges and the relationships between rotation curves with small root mean square deviations.

curve	**range**	**mean**
leIndexMCPx	0.20	0.5
leIndexMCPz	0.42	-3.8
leIndexDIPy	0.12	11.9
leMiddleMCPx	0.32	-0.7
leMiddleMCPz	0.95	-6.3
leMiddleDIPy	0.007	13.9
leRingDIPy	0.001	15.8
lePinkyDIPy	0.94	-0.5
riIndexDIPy	0.08	-8.0
riMiddleDIPy	0.08	-11.6
riPinkyDIPy	0.31	-16.4

Table D.1: Rotation curves with a range of less than 1.

curve	**range**	**mean**
leThumbCMCx	4.4	45.5
leRingMCPx	1.7	-1.8
leRingMCPz	2.2	-15.8
riIndexMCPx	4.8	1.2
riMiddleMCPx	4.7	-49.5
riRingMCPx	3.9	-0.3
riRingDIPy	1.5	-11.8
riPinkyPIPy	4.4	-12.2

Table D.2: Rotation curves with a range between 1 and 5.

dependent curve	**controlling curve**	**rmsd**
leRingMCPx	leRingMCPy	$\text{rmsd}_2 = 0.25$
leRingMCPz	leRingMCPy	$\text{rmsd}_2 = 0.19$
lePinkyMCPx	lePinkyMCPy	$\text{rmsd}_2 = 0.18$
lePinkyMCPz	lePinkyMCPy	$\text{rmsd}_2 = 0.21$
riThumbMCPz	riThumbIPz	$\text{rmsd}_1 = 0.17$
riIndexMCPx	riPinkyMCPy	$\text{rmsd}_2 = 0.18$
riMiddleMCPy	riPinkyMCPy	$\text{rmsd}_1 = 0.19$
riRingMCPx	riPinkyMCPz	$\text{rmsd}_2 = 0.22$
riRingMCPy	riPinkyMCPy	$\text{rmsd}_1 = 0.13$
riRingMCPz	riPinkyMCPy	$\text{rmsd}_1 = 0.13$
riRingDIPy	riRingPIPy	$\text{rmsd}_1 = 0.29$
riPinkyMCPx	riPinkyMCPz	$\text{rmsd}_2 = 0.25$

Table D.3: Relationships of rotation curves when all root mean square deviations of less than 0.3 are taken into account.

dependent curve	**controlling curve**	**rmsd**
leThumbCMCz	leThumbCMCy	$\text{rmsd}_2 = 0.46$
leIndexPIPy	leMiddlePIPy	$\text{rmsd}_1 = 0.50$
leRingMCPy	leMiddleMCPy	$\text{rmsd}_1 = 0.39$
leRingPIPy	leMiddlePIPy	$\text{rmsd}_1 = 0.45$
riThumbMCPz	riPinkyMCPy	$\text{rmsd}_1 = 0.51$
riThumbIPz	riPinkyMCPy	$\text{rmsd}_1 = 0.46$
riIndexMCPy	riPinkyMCPy	$\text{rmsd}_1 = 0.49$
riMiddlePIPy	riPinkyMCPy	$\text{rmsd}_1 = 0.35$
riRingPIPy	riPinkyMCPy	$\text{rmsd}_1 = 0.63$
riPinkyMCPx	riPinkyMCPy	$\text{rmsd}_2 = 0.56$
riPinkyMCPz	riPinkyMCPy	$\text{rmsd}_1 = 0.39$

Table D.4: Additional relationships of rotation curves when all root mean square deviations of less than 0.5 are taken into account.

Bibliography

[ADGY04] Atkinson A. P., Dittrich W. H., Gemmell A. J., Young A. W.: Emotion perception from dynamic and static body expressions in point-light and full-light displays. *Perception 33*, 6 (2004), 717–746.

[AHS03] Albrecht I., Haber J., Seidel H.-P.: Construction and animation of anatomically based human hand models. In *SCA '03: Proceedings of the 2003 ACM SIGGRAPH/Eurographics Symposium on Computer Animation* (2003), pp. 98–109.

[AN99] Aydin Y., Nakajima M.: Database guided computer animation of human grasping using forward and inverse kinematics. *Computers & Graphics 23*, 1 (1999), 145–154.

[Ari06] Arikan O.: Compression of motion capture databases. *ACM Transactions on Graphics 25*, 3 (2006), 890–897.

[AS03] Athitsos V., Sclaroff S.: Estimating 3d hand pose from a cluttered image. In *IEEE Computer Vision and Pattern Recognition (CVPR)* (June 2003), vol. 2, pp. 432–439.

[Aut11] Autodesk: 3ds Max website, http://usa.autodesk.com/3ds-max, February 2011.

[BHFSS06] Breakwell G. M., Hammond S., Fife-Schaw C., Smith J. A. (Eds.): *Research Methods in Psychology*, 3rd ed. SAGE Publications, 2006.

[BK00] Bicchi A., Kumar V.: Robotic grasping and contact: A review. In *ICRA* (2000), vol. 1, pp. 348–353.

[BRRP97] Bodenheimer B., Rose C., Rosenthal S., Pella J.: The process of motion capture: Dealing with the data. In *Computer Animation and*

Simulation '97 (Wien, 1997), Thalmann D., van de Panne M., (Eds.), Eurographics Animation Workshop, Springer-Verlag, pp. 3–18.

[BSH99] Bodenheimer B., Shleyfman A. V., Hodgins J. K.: The effects of noise on the perception of animated human running. In *Computer Animation and Simulation '99* (sep 1999), Magnenat-Thalmann N., Thalmann D., (Eds.), Springer-Verlag, Wien, pp. 53–63. Eurographics Animation Workshop.

[BSP*04] Barbič J., Safonova A., Pan J.-Y., Faloutsos C., Hodgins J. K., Pollard N. S.: Segmenting motion capture data into distinct behaviors. In *GI '04: Proceedings of Graphics Interface 2004* (2004), pp. 185–194.

[BW95] Bruderlin A., Williams L.: Motion signal processing. In *SIGGRAPH '95: Proceedings of the 22nd annual conference on Computer graphics and interactive techniques* (1995), pp. 97–104.

[BZ04] Braido P., Zhang X.: Quantitative analysis of finger motion coordination in hand manipulative and gestic acts. *Human Movement Science 22*, 6 (2004), 661–678.

[CFOC05] Chaminade T., Franklin D. W., Oztop E., Cheng G.: Motor interference between humans and humanoid robots: Effect of biological and artificial motion. *Proceedings of the 4th International Conference on Development and Learning, 2005* (2005), 96–101.

[CGA07] Ciocarlie M., Goldfeder C., Allen P.: Dimensionality reduction for hand-independent dexterous robotic grasping. In *Intelligent Robots and Systems, 2007* (2007), pp. 3270–3275.

[CHK07] Chaminade T., Hodgins J., Kawato M.: Anthropomorphism influences perception of computer-animated characters' actions. *Social Cognitive and Affective Neuroscience 2*, 3 (2007), 206–216.

[CK77] Cutting J. E., Kozlowski L. T.: Recognizing friends by their walk: Gait perception without familiarity cues. *Bulletin of the Psychonomic Society 9*, 5 (1977), 353–356.

[Cou04] Coulson M.: Attributing emotion to static body postures: Recognition accuracy, confusions, and viewpoint dependence. *Journal of Nonverbal Behavior 28*, 2 (2004), 117–139.

[CPMX07] Chang L. Y., Pollard N., Mitchell T., Xing E. P.: Feature selection for grasp recognition from optical markers. In *Proceedings of the 2007 IEEE/RSJ Intl. Conference on Intelligent Robots and Systems (IROS 2007)* (October 2007), pp. 2944 – 2950.

[CST*10] Carter E., Sharan L., Trutoiu L., Matthews I., Hodgins J. K.: Perceptually motivated guidelines for voice synchronization in film. *ACM Transactions on Applied Perception (TAP) 7* (July 2010), 23:1–23:12.

[CVB01] Cassell J., Vilhjálmsson H. H., Bickmore T.: Beat: the behavior expression animation toolkit. In *Proceedings of SIGGRAPH 2001* (2001), pp. 477–486.

[Cyb11] CyberGlove Systems: CyberGlove II product overview website, http://www.cyberglovesystems.com/products/cyberglove-ii/overview, February 2011.

[DGT11] DGTech Engineering Solutions: DG5 VHand description website, http://www.dg-tech.it/vhand/eng/index.html, February 2011.

[DSD08] Dipietro L., Sabatini A., Dario P.: A survey of glove-based systems and their applications. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on 38*, 4 (July 2008), 461–482.

[EBN*07] Erol A., Bebis G., Nicolescu M., Boyle R. D., Twombly X.: Vision-based hand pose estimation: A review. *Computer Vision Image Understanding 108*, 1-2 (2007), 52–73.

[ESGP06] El-Sawah A., Georganas N., Petriu E. M.: Finger inverse kinematics using error model analysis for gesture enabled navigation in virtual environments. In *IEEE International Workshop on Haptic Audio Visual Environments and their Applications* (November 2006), pp. 34–39.

[Fec60] Fechner G. T.: *Elemente der Psychophysik. Erster Theil.* Breitkopf und Härtel, 1860.

[Fif11] Fifth Dimension Technologies: 5DT Data Glove product description website, http://www.5dt.com/products/pdataglovemri.html, February 2011.

[Fil11] Filmsite: Filmsite website "Greatest Box-Office Bombs, Disasters and Flops", http://www.filmsite.org/greatestflops21.html, written by Tim Dirks, February 2011.

[FsM08] Feil-seifer D., Matarić M. J.: Toward socially assistive robotics for augmenting interventions for children with autism spectrum disorders. In *International Symposium on Experimental Robotics* (2008).

[Gd99] Greene J., d'Oliviera M.: *Learning to use statistical tests in psychology*, 2nd ed. Open University Press, 1999.

[Gel08] Geller T.: Overcoming the uncanny valley. *IEEE Computer Graphics and Applications 28*, 4 (2008), 11–17.

[Ges97] Gescheider G. A. (Ed.): *Psychophysics: the fundamentals*, 3rd ed. Lawrence Erlbaum Associates, 1997.

[GM08] Gibet S., Marteau P.-F.: Analysis of human motion, based on the reduction of multidimensional captured data – application to hand gesture compression, segmentation and synthesis. In *Articulated Motion and Deformable Objects* (2008), vol. 5098/2008, pp. 72–81.

[HFP*00] Herda L., Fua P., Plänkers R., Boulic R., Thalmann D.: Skeleton-based motion capture for robust reconstruction of human motion. In *CA '00: Proceedings of the Computer Animation* (2000), p. 77.

[HJH*04] Han S., Jiang Y., Humphreys G. W., Zhou T., Cai P.: Distinct neural substrates for the perception of real and virtual visual worlds. *NeuroImage 24*, 3 (2004), 928–935.

[HKG06] Heck R., Kovar L., Gleicher M.: Splicing upper-body actions with locomotion. *Computer Graphics Forum (Proceedings of Eurographics) 25*, 3 (2006), 459–466.

[HMP08] Ho C.-C., MacDorman K. F., Pramono Z. A. D. D.: Human emotion and the uncanny valley: a GLM, MDS, and Isomap analysis of robot

video ratings. In *HRI '08: Proceedings of the 3rd ACM/IEEE international conference on Human robot interaction* (2008), pp. 169–176.

[HOP*05] Hanson D., Olney A., Prilliman S., Mathews E., Zielke M., Hammons D., Fernandez R., Stephanou H.: Upending the uncanny valley. In *AAAI'05: Proceedings of the 20th national conference on Artificial intelligence* (2005), AAAI Press, pp. 1728–1729.

[HOT97] Hodgins J. K., O'Brien J. F., Tumblin J.: Do geometric models affect judgments of human motion? In *Graphics Interface '97* (1997), Davis W. A., Mantei M., Klassen R. V., (Eds.), Canadian Human-Computer Communications Society, pp. 17–25.

[HOT98] Hodgins J. K., O'Brien J. F., Tumblin J.: Perception of human motion with different geometric models. *IEEE Transactions on Visualization and Computer Graphics 4*, 4 (1998), 307–316.

[How08] Howell D. C.: *Fundamental Statistics for the Behavioral Sciences*, 6th ed. Thomson Wadsworth, 2008.

[HRvdP04] Harrison J., Rensink R. A., van de Panne M.: Obscuring length changes during animated motion. *ACM Transactions on Graphics 23*, 3 (2004), 569–573.

[HSD05] Hornung A., Sar-Dessai S.: Self-calibrating optical motion tracking for articulated bodies. In *VR '05: Proceedings of the 2005 IEEE Conference 2005 on Virtual Reality* (2005), pp. 75–82.

[HWBO95] Hodgins J. K., Wooten W. L., Brogan D. C., O'Brien J. F.: Animating human athletics. In *SIGGRAPH '95: Proceedings of the 22nd annual conference on Computer graphics and interactive techniques* (New York, NY, USA, 1995), ACM, pp. 71–78.

[IF04] Ikemoto L., Forsyth D. A.: Enriching a motion collection by transplanting limbs. In *ACM SIGGRAPH/Eurographics Symposium on Computer Animation* (2004), pp. 99–108.

[Ima11] Image Metrics: Facial animation software website, http://www.image-metrics.com, February 2011.

[Ish06] Ishiguro H.: Interactive humanoids and androids as ideal interfaces for humans. In *IUI '06: Proceedings of the 11th international conference on Intelligent user interfaces* (2006), pp. 2–9.

[Joh73] Johansson G.: Visual perception of biological motion and a model for its analysis. *Perception and Psychophysics 14*, 2 (1973), 201–211.

[Joh76] Johansson G.: Spatio-temporal differentiation and integration in visual motion perception. *Psychological Research 38* (1976), 379–393.

[JT05] Johnson K. L., Tassinary L. G.: Perceiving sex directly and indirectly: Meaning in motion and morphology. *Psychological Science 16*, 11 (2005), 890–897.

[JT07] Johnson K. L., Tassinary L. G.: Compatibility of basic social perceptions determines perceived attractiveness. *Proceedings of the National Academy of Sciences 104*, 12 (2007), 5246–5251.

[KC77] Kozlowski L. T., Cutting J. E.: Recognizing the sex of a walker from a dynamic point-light display. *Perception and Psychophysics 21*, 6 (1977), 575–580.

[KC78] Kozlowski L. T., Cutting J. E.: Recognizing the gender of walkers from point-lights mounted on ankles: Some second thoughts. *Perception and Psychophysics 23*, 5 (1978), 459.

[Keh07] Kehr D.: Duplicate motion, then capture emotion. *The New York Times* (2007). http://www.nytimes.com/2007/11/18/movies/18kehr.html, retrieved in August 2010.

[KGP02] Kovar L., Gleicher M., Pighin F.: Motion graphs. *ACM Transactions on Graphics 21*, 3 (2002), 473–482.

[KHW95] Kessler G. D., Hodges L. F., Walker N.: Evaluation of the CyberGlove as a whole-hand input device. *ACM Transactions on Computer-Human Interaction 2*, 4 (1995), 263–283.

[KNY05] Kozima H., Nakagawa C., Yasuda Y.: Interactive robots for communication-care: a case-study in autism therapy. *IEEE International*

Workshop on Robot and Human Interactive Communication (2005), 341–346.

[KOF05] Kirk A. G., O'Brien J. F., Forsyth D. A.: Skeletal parameter estimation from optical motion capture data. In *CVPR '05: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Volume 2* (2005), pp. 782–788.

[KP06] Kry P. G., Pai D. K.: Interaction capture and synthesis. *ACM Transactions on Graphics 25*, 3 (2006), 872–880.

[KP10] Krosnick J. A., Presser S.: Question and questionnaire design. In *Handbook of Survey Research*, Marsden P. V., Wright J. D., (Eds.), 2nd ed. Emerald Group, 2010.

[KPB03] Kilner J., Paulignan Y., Blakemore S.: An interference effect of observed biological movement on action. *Current Biology 13*, 6 (2003), 522–525.

[KW08] Kitagawa M., Windsor B.: *MoCap for Artists: Workflow and Techniques for Motion Capture*. Focal Press, 2008.

[KZK04] Kahlesz F., Zachmann G., Klein R.: Visual-fidelity dataglove calibration. In *CGI '04: Proceedings of the Computer Graphics International* (2004), pp. 403–410.

[LCG03] Lattin J. M., Carroll J. D., Green P. E.: *Analyzing Multivariate Data*. Brooks/Cole – Thomson Learning, 2003.

[LCR*02] Lee J., Chai J., Reitsma P. S. A., Hodgins J. K., Pollard N. S.: Interactive control of avatars animated with human motion data. *ACM Transactions on Graphics 21*, 3 (2002), 491–500.

[LFP07] Li Y., Fu J. L., Pollard N. S.: Data-driven grasp synthesis using shape matching and task-based pruning. *IEEE Transactions on Visualization and Computer Graphics 13*, 4 (2007), 732–747.

[Liu08] Liu C. K.: Synthesis of interactive hand manipulation. In *SCA '08: Proceedings of the 2008 ACM SIGGRAPH/Eurographics Symposium on Computer Animation* (2008), pp. 163–171.

[Liu09] Liu K. C.: Dextrous manipulation from a grasping pose. *ACM Transactions on Graphics 28*, 3 (August 2009), 3:1–3:6.

[Mac05] MacDorman K. F.: Mortality salience and the uncanny valley. In *Proceedings of the 5th IEEE-RAS International Conference on Humanoid Robots* (2005), pp. 399–405.

[Mac06] MacDorman K. F.: Subjective ratings of robot video clips for human likeness, familiarity, and eeriness: An exploration of the uncanny valley. In *IUI '06: Proceedings of the 11th international conference on Intelligent user interfaces* (2006), pp. 26–29.

[McN92] McNeill D.: *Hand and Mind: what gestures reveal about thought.* The University of Chicago Press, 1992.

[MEDO09] McDonnell R., Ennis C., Dobbyn S., O'Sullivan C.: Talking bodies: Sensitivity to desynchronization of conversations. In *APGV '09: Proceedings of the 6th Symposium on Applied Perception in Graphics and Visualization* (2009).

[Men99] Menache A.: *Understanding Motion Capture for Computer Animation and Video Games.* Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1999.

[MI06] MacDorman K., Ishiguro H.: The uncanny advantage of using androids in cognitive and social science research. *Interaction Studies 7*, 3 (2006), 297–337.

[MIWFPB02] Meehan M., Insko B., Whitton M., Frederick P. Brooks J.: Physiological measures of presence in stressful virtual environments. *ACM Transactions on Graphics 21*, 3 (2002), 645–652.

[MLH*09] McDonnell R., Larkin M., Hernández B., Rudomin I., O'Sullivan C.: Eye-catching crowds: saliency based selective variation. *ACM Transactions on Graphics 28*, 3 (2009), 1–10.

[MM94] Mather G., Murdoch L.: Gender discrimination in biological motion displays based on dynamic cues. *Proceedings of the Royal Society of London, Series B 258* (1994), 273–279.

[MNO07] McDonnell R., Newell F., O'Sullivan C.: Smooth movers: perceptually guided human motion simulation. In *ACM SIGGRAPH/Eurographics Symp. on Comp. Animation* (2007), pp. 259–269.

[Mor70] Mori M.: The uncanny valley. *Energy 7*, 4 (1970), 33–35.

[Mov11] Mova: Mova website, http://www.mova.com, February 2011.

[MZF06] Majkowska A., Zordan V. B., Faloutsos P.: Automatic splicing for hand and body animations. In *SCA '06: Proceedings of the 2006 ACM SIGGRAPH/Eurographics Symposium on Computer Animation* (2006), pp. 309–316.

[Nap80] Napier J.: *Hands*. New York: Pantheon Books, Princeton, NJ, 1980.

[Nat11] NaturalMotion: Morpheme 3 product website, http://www.naturalmotion.com/morpheme.htm, February 2011.

[OBBH00] O'Brien J. F., Bodenheimer R., Brostow G., Hodgins J. K.: Automatic joint parameter estimation from magnetic motion capture data. In *Graphics Interface* (2000), pp. 53 – 60.

[OFH08] Onuma K., Faloutsos C., Hodgins J. K.: FMDistance: A fast and effective distance function for motion capture. In *Eurographics Short Papers* (2008).

[OGR11] OGRE: Open-Oriented Graphics Rendering Engine website, http://www.ogre3d.org, February 2011.

[Ope11] OpenAL: Open Audio Library website, http://connect.creativelabs.com/openal, February 2011.

[Par07] Parent R.: *Computer Animation: Algorithms & Techniques*, 2nd ed. Morgan Kaufmann, 2007.

[Pel05] Pelachaud C.: Multimodal expressive embodied conversational agents. In *MULTIMEDIA '05: Proceedings of the 13th annual ACM international conference on Multimedia* (2005), pp. 683–689.

[PFB*01] Perani D., Fazio F., Borghese N. A., Tettamanti M., Ferrari S., Decety J., Gilardi M. C.: Different brain correlates for watching real and virtual hand actions. *NeuroImage 14* (2001), 749–758.

[PFS06] Palastanga N., Field D., Soames R.: *Anatomy and Human Movement – Structure and Function*, 5 ed. Butterworth Heinemann Elsevier, 2006.

[PH06] Park S. I., Hodgins J. K.: Capturing and animating skin deformation in human motion. In *SIGGRAPH '06: ACM SIGGRAPH 2006 Papers* (2006), pp. 881–889.

[PMM*03] Pelphrey K. A., Mitchell T. V., McKeown M. J., Goldstein J., Allison T., McCarthy G.: Brain activity evoked by the perception of human walking: controlling for meaningful coherent motion. *The Journal of Neuroscience 23*, 17 (2003), 6819–6825.

[PSH97] Pavlovic V. I., Sharma R., Huang T. S.: Visual interpretation of hand gestures for human-computer interaction: A review. *IEEE Transactions on Pattern Analysis Machine Intelligence 19*, 7 (1997), 677–695.

[PZ05] Pollard N. S., Zordan V. B.: Physically based grasping control from example. In *SCA '05: Proceedings of the 2005 ACM SIGGRAPH/Eurographics symposium on Computer animation* (2005), pp. 311–318.

[RAP08] Reitsma P. S. A., Andrews J., Pollard N. S.: Effect of character animacy and preparatory motion on perceptual magnitude of errors in ballistic motion. *Computer Graphics Forum 27*, 2 (apr 2008), 201–210.

[RG91] Rijpkema H., Girard M.: Computer animation of knowledge-based human grasping. *SIGGRAPH Comput. Graph. 25*, 4 (1991), 339–348.

[Rot11a] Rotten Tomatoes: Movie reviews website, "Final Fantasy – The Spirits Within (2001)", http://www.rottentomatoes.com/m/final_fantasy_the_spirits_within, February 2011.

[Rot11b] Rotten Tomatoes: Movie reviews website, "The Polar Express (2004)", http://www.rottentomatoes.com/m/polar_express, February 2011.

[RP03] Reitsma P. S. A., Pollard N. S.: Perceptual metrics for character animation: sensitivity to errors in ballistic motion. *ACM Transactions on Graphics 22*, 3 (July 2003), 537–542.

[SFS98] Santello M., Flanders M., Soechting J. F.: Postural hand synergies for tool use. *The Journal of Neuroscience 18*, 23 (1998), 10105–10115.

[SG07] Saris W. E., Gallhofer I. N.: *Design, Evaluation, and Analysis of Questionnaires for Survey Research*. John Wiley & Sons, 2007.

[SH05] Safonova A., Hodgins J. K.: Analyzing the physical correctness of interpolated human motion. In *SCA '05: Proceedings of the 2005 ACM SIGGRAPH/Eurographics symposium on Computer animation* (2005), pp. 171–180.

[SHP04] Safonova A., Hodgins J. K., Pollard N. S.: Synthesizing physically realistic human motion in low-dimensional, behavior-specific spaces. In *SIGGRAPH '04: ACM SIGGRAPH 2004 Papers* (2004), pp. 514–521.

[SKP08] Sueda S., Kaufman A., Pai D. K.: Musculotendon simulation for hand animation. *ACM Transactions on Graphics 27*, 3 (Aug. 2008), 3:1–3:8.

[SKY07] Shum H. P. H., Komura T., Yamazaki S.: Simulating competitive interactions using singly captured motions. In *VRST '07: Proceedings of the 2007 ACM symposium on Virtual reality software and technology* (2007), pp. 65–72.

[SMII06] Shimada M., Minato T., Itakura S., Ishiguro H.: Evaluation of android using unconscious recognition. In *6th IEEE-RAS International Conference on Humanoid Robots* (2006), pp. 157–162.

[Ste46] Stevens S. S.: On the theory of scales of measurement. *Science 103*, 2684 (1946), 677–680.

[SWY07] Schneider E., Wand Y., Yang S.: Exploring the uncanny valley with japanese video game characters. In *Situated Play of DiGRA Conference (Digital Games Research Association* (2007), pp. 546–549.

[SZ94] Sturman D. J., Zeltzer D.: A survey of glove-based input. *IEEE Computer Graphics and Applications 14*, 1 (1994), 30–39.

[TG09a] Tinwell A., Grimshaw M.: Bridging the uncanny: an impossible traverse? In *MindTrek '09: Proceedings of the 13th International MindTrek Conference: Everyday Life in the Ubiquitous Era* (2009), pp. 66–73.

[TG09b] Tinwell A., Grimshaw M.: Survival horror games - an uncanny modality. In *Games Computing and Creative Technologies* (2009).

[TJ81] Thomas F., Johnston O.: *Disney Animation: The Illusion of Life.* Abbeville Press, 1981.

[Tob08] Tobii Support: *Tobii Studio^{TM}1.2 User Manual*, June 2008.

[VGK02] Vlachos M., Gunopoulos D., Kollios G.: Discovering similar multidimensional trajectories. In *ICDE '02: Proceedings of the 18th International Conference on Data Engineering* (2002), IEEE Computer Society, p. 673.

[Vic11] Vicon: Motion capture system website, http://www.vicon.com, February 2011.

[Vir11] Virtual Realities: Data gloves website, http://www.vrealities.com/glove.html, February 2011.

[VS07] Vaz M. C., Starkey S.: *The Art of Beowulf.* Chronicle Books, San Francisco, California, 2007.

[Wal98] Wallbott H. G.: Bodily expression of emotion. *European Journal of Social Psychology 28*, 6 (1998), 879–896.

[WB04] Wang J., Bodenheimer B.: Computing the duration of motion transitions: an empirical approach. In *SCA '04: Proceedings of the 2004 ACM SIGGRAPH/Eurographics symposium on Computer animation* (2004), pp. 335–344.

[Wil01] Williams R.: *The Animator's Survival Kit.* Faber and Faber, London – New York, 2001.

[WP95] Witkin A., Popovic Z.: Motion warping. In *SIGGRAPH '95: Proceedings of the 22nd annual conference on Computer graphics and interactive techniques* (1995), pp. 105–108.

[WP09] Wang R. Y., Popović J.: Real-time hand-tracking with a color glove. *ACM Transactions on Graphics 28*, 3 (2009), 1–8.

[WWXZ06] Wen G., Wang Z., Xia S., Zhu D.: From motion capture data to character animation. In *VRST '06: Proceedings of the ACM symposium on Virtual reality software and technology* (2006), pp. 165–168.