Helsinki University of Technology
Publications in Telecommunications Software and Multimedia
Teknillisen korkeakoulun tietoliikenneohjelmistojen ja multimedian julkaisuja
Espoo 2007 TML-A18

NOVEL APPLICATIONS OF REAL-TIME AUDIOVISUAL SIGNAL PROCESSING TECHNOLOGY FOR ART AND SPORTS EDUCATION AND ENTERTAINMENT

Perttu Hämäläinen

Dissertation for the degree of Doctor of Science in Technology to be presented with due permission of the Department of Computer Science and Engineering, for public examination and debate in Auditorium T2 at Helsinki University of Technology (Espoo, Finland) on the 30th of March, 2007, at 12 noon.

Helsinki University of Technology
Department of Computer Science and Engineering
Telecommunications Software and Multimedia Laboratory

Teknillinen korkeakoulu
Tietotekniikan osasto
Tietoliikenneohjelmistojen ja multimedian laboratorio

Distribution:
Helsinki University of Technology
Telecommunications Software and Multimedia Laboratory
P.O.Box 5400
FIN-02015 TKK
URL: http://www.tml.tkk.fi
Tel. +358-9-451 2870
Fax. +358-9-451 5014

ISBN 978-951-22-8704-8
ISBN 978-951-22-8705-5 (PDF)
ISSN 1456-7911
ISSN 1455-9722 (PDF)

URL: http://lib.tkk.fi/Diss/2007/isbn9789512287055/

Multiprint Oy/Otamedia
Espoo 2007

ABSTRACT

Author Perttu Hämäläinen
Title Novel applications of real-time audiovisual signal processing technology for art and sports education and entertainment

This thesis explores the possibilities of real-time audiovisual signal processing in the context of human-computer interfaces. Novel embodied interaction applications are presented, including two computer games controlled by body movements and voice, two computer games played by singing, software and an installation for creating clay animation, and interactive video "magic" mirrors for martial arts training. The common denominators for the case studies are 1) the use of cameras and microphones as the input devices, 2) processing and analyzing the audiovisual input to generate visual feedback that can make it easier for the users to detect and correct errors in their performance, and 3) a holistic approach to interaction design, including not only the user and a computer, but the context of interaction as well, e.g., the social setting and the differences between the real and the virtual environments.

In edutainment (education and entertainment) software, the interface sets the limits for the user's actions. The thesis demonstrates how embodied interaction can widen the scope of edutainment to encompass new skills, in this case martial arts, singing, and modelling and animating clay. The thesis also contributes to computer vision technology in the form of practical vision systems that interpret low-level image features according to the application context. For example, in Kick Ass Kung-Fu, an embodied martial arts game, several simultaneous users can fight against computer-generated virtual enemies. The players can wield weapons such as swords, and no specific clothing or markers are needed.

Keywords computer vision, interaction design, signal processing, mixed reality, sports, animation, music, computer games

TIIVISTELMÄ

Tekijä Perttu Hämäläinen
Otsikko Uusia reaaliaikaiseen kuvan- ja äänenkäsittelyteknologiaan perustuvia järjestelmiä taiteen ja liikunnan harjoitteluun

Käyttöliittymäteknologia, esimerkiksi syöttölaite, rajoittaa usein käyttäjän toimintaa ohjelmistosovelluksissa. Kehollisen vuorovaikutuksen teknologiat ja käytännöt mahdollistavat yhä useampien taitojen, esimerkiksi liikunnan ja musiikin harjoittelemisen digitaalisissa ympäristöissä.

Väitöskirjatyössä kehitettiin uusia kehollisia oppiviihdesovelluksia ja käyttöliittymäprototyyppejä, mukaanlukien kaksi keholla ja äänellä pelattavaa tietokonepeliä, kaksi laulamalla pelattavaa tietokonepeliä, vaha-animaatio-ohjelmisto ja –installaatio sekä vuorovaikutteisia videopeilejä kamppailulajien ja akrobatian harjoitteluun. Projekteja yhdistää se, että 1) tietokoneen syöttölaitteina käytetään kameraa ja/tai mikrofonia, 2) käyttäjälle tuotetaan virheiden havaitsemista ja korjaamista avustavaa visuaalista palautetta kuva- ja äänisyötteen analyysiin perustuen ja 3) vuorovaikutussuunnittelua pyritään lähestymään kokonaisvaltaisesti siten, että käyttäjän ja teknologian lisäksi huomioidaan myös vuorovaikutustilanne, esimerkiksi sosiaalinen ympäristö ja tosi- ja virtuaalimaailmojen väliset erot.

Väitöskirjatyössä kehitettiin myös uutta tietokonenäköteknologiaa käytettävyyssuunnittelun vaatimusten mukaisesti. Esim. kamppailupeli Kick Ass Kung-Fu:ssa useat pelaajat voivat taistella yhteistyössä tietokonevastustajia vastaan. Pelin tietokonenäkö ei aseta rajoituksia vaatetuksen tai käytettävien aseiden suhteen.

Avainsanat tietokonenäkö, vuorovaikutussuunnittelu, signaalinkäsittely, liikunta, animaatio, musiikki, tietokonepelit

ACKNOWLEDGEMENTS

First of all, I would like to thank my colleagues and co-authors Mikko Lindholm, Ari Nykänen, Johanna Höysniemi, Laura Turkki, Teppo Rouvi, Teemu Mäki-Patola, Tommi Ilmonen, Ville Pulkki, Matti Airas, Sonja Kangas, and Johanna Meltaus. The various projects described in this thesis are the result of a lot of collaborative effort. Besides the authors, the projects owe a lot to the people who participated in the user studies, e.g., Minna Nevanlinna and Anne-Mari Lindholm, whose insights were very helpful in designing Animaatiokone, and Riikka Pelo, both an excellent voice actor and mother to the wonderful Oskari and Eemeli who evaluated the first prototype of QuiQui's Giant Bounce.

Considering the written part of the thesis, the quality was greatly improved by the suggestions made by the pre-examiners Kari-Jouko Räihä and Ernest Edmonds. I also thank Jaakko Kulonpalo for proofreading.

The thesis experiments with digitally enhanced learning and practicing of arts and sports. I would not have chosen the topic, and I would not understand a thing about motor learning and all things embodied, had I not practiced arts and sports myself. I thank my teachers Jon Sundwall and Hans-Kristian Sjöholm (TaeKwonDo), Peter Nystén (Karate), Guy Windsor (medieval swordmanship), Kurt Berger and Markus Vainio-Mattila (Capoeira and acrobatics), Anita Isomettä (modern dance), Antero Kuisma ja Suojatyöpaikka (Henri Penttinen, Petri Naukkarinen, Janne Toivanen, Jussi Koski, and Kristian Sahlstedt), and all my fellow trainees, especially Temppukerho (Antti Koivunen, Riku Ruokolahti, and Aleksi Pihkanen).

Although most of my closest colleagues work outside Helsinki University of Technology (HUT), I have been using the Telecommunications Software and Multimedia Laboratory (TML) as my base of operations between 2002 and 2006. I am grateful to Tapio Takala and Lauri Savioja for supervision and guidance. I am also thankful to the entire EVE-team at TML, and Petri Vuorimaa, the head of the laboratory. Thanks to Eeva Kykkänen, Liisa Hirvisalo, Ilpo Lahtinen, Seppo Äyräväinen, Esa Virtanen, and Esa Heikkinen for providing support and for all the gear I constantly bugged you for. Thanks to Wille Mäkelä, Tapio Lokki, Samuli Laine, Janne Kontkanen, Jaakko Lehtinen, Ari Silvennoinen, and Timo Aila for the interesting conversations. The 3Dr group must also be acknowledged for the atmosphere – for a laboratory, the smell of good coffee is what the smell of *pulla* is for a safe home and a happy family =)

Before TML, I worked at Elmorex Ltd, where we developed the singing games described in this thesis. I thank the entire Elmorex team, of which I have not yet mentioned Jyrki Kohonen, Erno Soinila, Niina Saarelainen, and Tero Tolonen. My tools are digital, and Dr. Tolonen provided me with my first "hammer". He developed the original version of the pitch detection technology used for the Elmorex games.

After submitting the thesis for pre-examination, I moved from HUT to the Virtual Air Guitar Company. I want to thank Teemu Mäki-Patola, Aki Kanerva, Juha Laitinen, Tommi Tykkälä, Markus Eräpolku, Pirjo Kekäläinen-Torvinen and Taina Myöhänen for the inspiring and fun environment. It could have been a lot harder to fire up the computer after a full day's work and start adding the finishing touches to the thesis.

Finally, an extra special thank-you-so-much-hug-and-kisses to Helka for keeping me sane during the writing. Having you close to me anchored me to reality during the nights I feared I'd get lost inside my mind. My sanity has also been greatly supported by all the *tyypit* (they know who), and the countless hours of playing Liero and Soldat with Teemu, Aki, Magge, Juha and the graphics boys. Hail the creators of the games! Thanks to Pekka for the love affair counselling and all the experiences we shared in our brand new lives. Dear mom and dear sister Iippa, I hope I will get to share more time with you now that the thesis is finished.

My work has been funded by the Helsinki Graduate School on Computer Science and Engineering of the Academy of Finland. Additionally, I have been supported by YLE, Alfred Kordelin Foundation, University of Arts and Design Helsinki, Nokia Foundation, Jenny and Antti Wihuri foundation, KAUTE-foundation, and Tekniikan edistämissäätiö.

Helsinki, 21st of February 2007,

Perttu Hämäläinen

TABLE OF CONTENTS

LIST OF PUBLICATIONS

This thesis summarizes the following articles and publications, referred to as [P1]-[P6]:

[P1] Perttu Hämäläinen and Johanna Höysniemi. 2003. A Computer Vision and Hearing Based User Interface for a Computer Game for Children. In N. Carbonell, C. Stephanidis (Eds.): *User Interfaces for All,* LNCS 2615*,* 299–318, Springer-Verlag Berlin Heidelberg. First appeared in *Proceedings of the 7th ERCIM Workshop "User Interfaces for All"*, Paris, France, October 23-25, 2002

[P2] Perttu Hämäläinen, Mikko Lindholm, Ari Nykänen and Johanna Höysniemi. 2004. Animaatiokone - an Installation for Creating Clay Animation. *Proceedings of ACM Conference on Human Factors in Computing Systems (CHI 2004)*, Vienna, Austria, April 24-29, 2004, 17-24

[P3] Perttu Hämäläinen, Teemu Mäki-Patola, Ville Pulkki and Matti Airas. 2004. Musical Computer Games Played by Singing. *Proceedings of the 7th International Conference on Digital Audio Effects (DAFx'04)*, Naples, Italy, October 5-8, 2004, 367-371

[P4] Perttu Hämäläinen. 2004. Interactive Video Mirrors for Sports training. *Proceedings of the Third Nordic Conference on Human-Computer Interaction*, Tampere, Finland, October 23-27, 2004, 199-202

[P5] Perttu Hämäläinen, Johanna Höysniemi, Tommi Ilmonen, Mikko Lindholm and Ari Nykänen. 2005. Martial Arts in Artificial Reality. *Proceedings of ACM Conference on Human Factors in Computing Systems (CHI'2005)*, Portland, Oregon, April 2-7 2005, 781-790

[P6] Johanna Höysniemi and Perttu Hämäläinen. 2006. Children’s and Parents’ Perception of Full-Body Interaction and Violence in a Martial Arts Game. *GAIN: AIGA Journal of Business and Design*, March 7, 2006. First appeared in *Proceedings of the 2005 conference on Designing for User eXperience (DUX ‘05)*, San Francisco, CA, USA, November 3-5, 2005

AWARDS AND EXHIBITIONS

This thesis is multidisciplinary, situated at the intersection of human-computer interaction research, media art, and interaction design. In addition to the publications, the results of the thesis have been disseminated through the following festivals, exhibitions and competitions.

Awards

2004	Kick Ass Kung-Fu (Perttu Hämäläinen, Ari Nykänen, Mikko Lindholm) won the Games Platforms category in Europrix Top Talent multimedia innovation competition in Vienna, Austria.
	Tampere Mindtrek Grand Prix for Kick Ass Kung-Fu, Tampere, Finland
	Korjaamo Young Design Award for Animaatiokone (Mikko Lindholm, Perttu Hämäläinen, Ari Nykänen), Helsinki, Finland
2003	Animaatiokone won the Pikku Kakkonen (best multimedia for children) and non-commercial categories at Tampere Mindtrek competition, Finland.
	Prix Spécial du Jury for Animaatiokone in the international Prix Möbius multimedia competition in Athens, Greece.
2002	Kukakumma Muumaassa (Perttu Hämäläinen, Johanna Höysniemi, Teppo Rouvi, Laura Turkki) was one of the finalists of the Milia New Talent Competition in Cannes, France, February 2002.
2001	Kukakumma Muumaassa won the Pikku Kakkonen category in Tampere Mindtrek competition (best multimedia for children)

Exhibitions and festivals

2006	Kick Ass Kung-Fu and Animaatiokone at WIRED NextFest, New York, US
	Kukakumma Muumaassa at Ars Electronica Center, Linz, Austria
	Kick Ass Kung-Fu at Mal au Pixel, Paris, France
2005	Kick Ass Kung-Fu at Ultrasound festival, Huddersfield, UK
	An invited talk about Kick Ass Kung-Fu at Flimmer international film festival, Norrköping, Sweden

Animaatiokone at Japan Pop, Tennis Palace Art Museum, Helsinki, Finland

Kick Ass Kung-Fu at WIRED NextFest, Chicago, US

An invited talk about Kick Ass Kung-Fu at International symposium on Mixed Reality, RIXC Media Space, Riga, Latvia

Kick Ass Kung-Fu at Tennispalatsi, Helsinki, Finland

Kick Ass Kung-Fu at pixelACHE 2005 festival, Helsinki, Finland

Kick Ass Kung-Fu at Korjaamo gallery, Helsinki, Finland

2004 Kick Ass Kung-Fu at Nordic Exceptional Trendshop, Copenhagen, Denmark

Kick Ass Kung-Fu at Europrix Top Talent, Vienna, Austria

Kick Ass Kung-Fu at MindTrek, Tampere, Finland

Kick Ass Kung-Fu at Kiasma Theater, Helsinki, Finland

Animaatiokone at Korjaamo gallery, Helsinki, Finland

Animaatiokone at Animex festival, Middlesbrough, UK

Kukakumma Muumaassa at Musta tuntuu, Rauma Art Museum, Rauma, Finland

2003 Kukakumma Muumaassa at Game On exhibition, Tennis Palace Art Museum, Helsinki, Finland

Animaatiokone at Nordic Exceptional Trendshop, Copenhagen, Denmark

Animaatiokone at MindTrek, Tampere, Finland

Animaatiokone at Prix Möbius International, Athens, Greece

Animaatiokone at Annecy animation festival, Annecy, France

Animaatiokone at Tough Eye animation festival, Turku, Finland

Animaatiokone at Kunsthalle Lophem, Belgium

Animaatiokone at PixelAche: Kiasma in Helsinki, Finland, Gerschwin Hotel in New York, US, and SAT in Montreal, Canada

Animaatiokone at FF media festival, Rovaniemi, Finland

Animaatiokone at Tampere Film Festival, Tampere, Finland

Animaatiokone at Prix Möbius Preview Exhibition, Helsinki, Finland

2002 Kukakumma Muumaassa at Design Museo, Helsinki, Finland

Animaatiokone at Leffakansio, Helsinki, Finland

Animaatiokone at Kettupäivät, Helsinki, Finland

Animaatiokone at Animatricks, Helsinki, Finland

Animaatiokone at Mainonnan Viikko, Helsinki, Finland

Animaatiokone at Love & Anarchy film festival, Helsinki, Finland

1 INTRODUCTION

"Lihavat lapset saatava liikkeelle" (Obese children need exercise), states a headline in Helsingin Sanomat, a national Finnish newspaper on Sunday, 6th of March 2005 [41]. The article discusses the increasing percentage of overweight children and the role of information technology. The author of the article, Mikael Fogelholm, is not alone. Over the years, it has been a common concern in the media that computers and video games compete with sports as children's leisure activities. Many studies suggest that computers and video games promote a sedentary lifestyle and contribute to the increase of obesity [98][116].

Digitally promoted obesity, however, is only a part of a more profound problem. Computers have become part of our everyday lifestyle, but human-computer interaction (HCI) is still largely based on a simplified view of humans as sedentary beings with the brain to think and fingers to type with. Unless a more holistic, embodied view becomes prevalent, the *digitalization of every-day experiences* can deprive us of the physical aspects of life. Sports video games are an example of this, especially if they are not an addition but a replacement for the actual physical activity. This thesis, among many others, describes attempts to remedy this by developing human-computer interfaces that engage the user’s body and voice in the interaction.

There is a growing body of related research on novel human-computer interaction paradigms and technology. In a wider perspective, this thesis is a part of the age-old philosophical and scientific debate over the relationship of the human mind and body. Historically, influential religions and philosophers (e.g., Descartes) have promoted a dualistic view of humans as immortal minds inhabiting material, machine-like bodies (see, e.g., Bracken and Thomas [13], Robinson and Groves [99], and Wikipedia [19]). In light of modern science, the mind and the body seem inseparable: many aspects of human behaviour can be explained trough natural selection and genetic adaptation [121], neurological causes for mental disorders have been found [28], and there is evidence suggesting that physical exercise can promote mental health, e.g., by alleviating depression [88][122]. Related to HCI, Dourish has investigated human-computer interaction from an anti-Cartesian, phenomenological point of view [34]. His notion of embodied interaction, a unifying view of physical and social computing with focus on users as embodied agents, is something that could be used as an umbrella term for the developments described in this thesis.

1.1 Research questions

This thesis has the following main research questions:

- What kind of novel educational software can be created with help of real-time video and audio signal processing technology applied to human-computer interaction? In all the experiments described in this thesis, one doesn't use a computer with traditional control devices, such as a keyboard and a mouse. Instead, a camera and a microphone are used as the eyes and the ears of the computer, so that one can move and shout to interact with the software. This kind of perceptual and embodied interaction has been researched for decades, but with the decreasing cost of video and computer hardware, the range of real-life applications has increased dramatically during the past few years.
- If novel educational software can be developed, what is the role of the user interface and interface technology?

1.2 Research process

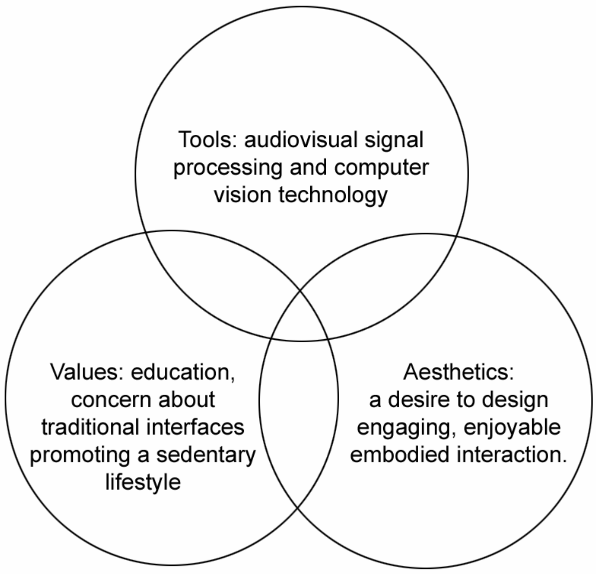

Figure 1. The research and design approach as a combination of aesthetics, values and tools.

The research questions were not clear at the beginning, but rather emerged from HCI experiments motivated by a combination of personal values and my scientific and artistic background. The research process can be considered an exploration of the space of possibilities in the centre of the Venn diagram in Figure 1. One of my main values is that I should be doing something useful, hence the educational theme of the thesis. My artistic background has given me a certain vision of the aesthetics of HCI, a desire to create engaging interactive experiences. Complementing this, my scientific background has given me interesting tools to explore: novel audiovisual signal processing and computer vision technology.

1.3 Scope: a collection of case studies

HCI literature is nowadays rife with studies of novel user interfaces and interaction design. The topic is approached from various angles, of which at least the following are relevant here:

- Case studies that propose and evaluate new ideas. This is mainly what this thesis is about. Case studies often contribute to the field by introducing and evolving reusable interface elements. For example, alternatives for the traditional drop-down menu have been developed [5], novel metaphors have been proposed in addition to the conventional thrash cans, folders etc. [128], and interface elements have been enhanced with animations and real-time transformations [6][21][124]. Case studies try out design methods and guidelines, and describe problems and possible solutions related to novel applications, ranging from utility software to art, entertainment and baffling innovations that people create simply because they can. Successful case studies also act as inspiration for researchers and designers. Series of case studies often provide insight on the properties of emerging classes of interfaces and applications, as in the case of the perceptive spaces and physically interactive story environments by Pinhanez *et al.* and Wren *et al.* [91][137]. As described in Chapter 4, the case studies in this thesis represent an emerging class of psychomotor edutainment.
- Methodology and guidelines, e.g., methods and tools for prototyping perceptual user interfaces [108], design principles for camera-based computer games [32], and guidelines for usability testing and collaborative design with children [35][51]. Novel methodology for usability evaluation and prototyping was also developed as part of the thesis work [57][58][59], but the details are beyond the scope of this thesis. The methodology is in a more prominent role in the doctoral thesis of my colleague Johanna Höysniemi.
- Technology for detecting and understanding user input, e.g., face tracking [14] and computer vision-based human motion capture in general [136][81]. This thesis describes some novel computer vision technology, although the process was not technology driven. Technology was developed only when necessary.
- Psychology, e.g., studies of the emotions of the users [118][119], and the effects of imperfect technology, such as latency, a common concern especially in musical user interfaces [40][42][103]. This is not my main area of expertise. Although some related observations are made in this thesis, I try to be cautious in drawing conclusions.

Chronologically, my work began with computer games played by singing into a microphone [P3]. In those games, the computer visualizes the pitch of the user's voice, which can help in singing in tune [55][130]. The role of computer-generated visual feedback in learning is the central theme for this thesis, although I only realized it in retrospect when writing publication [P3].

After the singing games, I continued by adding gestures as an input modality, resulting in Kukakumma Muumaassa (QuiQui's Giant Bounce), a sports adventure game for children that aimed to get children moving when playing computer games [P1]. Starting from a webcam-based animation tool originally developed for Kukakumma, I sidestepped into the realm of computer-aided animation. The result was Animaatiokone, an installation/machine for creating clay animation, where visual feedback again played a crucial role, for example, in making the animator aware of the amount of movement between consecutive frames of animation [P2]. Finally, I moved on to Kick Ass Kung-Fu [P4][P5][P6], a martial arts game for teenagers and adults, that recycles and refines the technology and concepts of the earlier projects, especially the utilization of visual feedback for motor learning.

1.4 Structure

The following treatment is organized so that Chapter 2 first reviews the role of feedback in learning and gives an overview of the case studies. To keep the overview compact, Chapter 2 focuses on concepts and interaction design, and technological issues are discussed in Chapter 3. Finally, Chapter 4 takes a retrospective look at the case studies as a whole and reflects upon them in the context of edutainment (education and entertainment) and educational simulations.

2 NOVEL FEEDBACK SYSTEMS FOR LEARNING AND PERFORMANCE

This thesis is a collection of HCI case studies that describe and evaluate novel applications for audiovisual signal processing technology. The unifying theme for the case studies is education, or more specifically, learning and practicing *skilled performances*, in this case athletic skills, singing, and creating and animating clay figures. Using the terminology of movement science, a *skill* denotes a capability that is developed through practice, in contrast to an *ability*, which is mostly determined genetically [104].

The main design principle is simple: A combination of screens, cameras and microphones is used to provide visual feedback that can help in analyzing and improving one's skills. It should be noted that in an interactive system, there is feedback by definition - human-computer interaction can be defined as dialogue between the user and the computer, a feedback loop of input and output [25].

2.1 Background

The role of feedback in learning and performance

There is several decades of research on the role of feedback in learning and performance. Overviews and a history of feedback research can be found, e.g., in papers by Bilodeau, Kluger and Denisi, Balzer *et al.*, and Newell [4][8][67][84]. Menickelli has produced a review of feedback literature as part of his PhD thesis [80]. Schmidt and Wrisberg also discuss the role of feedback in motor learning and performance [104].

Feedback, or feedback intervention (FI), means that people are provided information regarding their task performance [67]. Thorndike presented the initial theoretical arguments for the effectiveness of feedback in 1913 [67]. According to feedback intervention theory (FIT), *"behaviour is regulated by comparisons of feedback with goals or standards (and identification of gaps between the two)"* [67]. It has been asserted that *"feedback enables individuals to understand and improve their judgments, improve their expertise in the judgment task, and reduce commitment to incorrect judgment strategies"* [4].

There are various terms and acronyms for specific kinds of feedback, such as knowledge of results (KR) and knowledge of performance (KP), both of which are also referred to as outcome feedback (OFB) [4]. An example of KR is when a physical therapist says "now" when a patient's limb has become fully extended, whereas KP refers to movement characteristics, such as "your legs were slightly bent" [80]. However, KR and KP are also used in an ambiguous manner [67]. Bilodeau commented the diversity of feedback terminology as *"knowledge and feedback represent the core words, modified by other words such as results, performance, psychological, achievement, intrinsic, extrinsic, extra, supplementary, augmented, degraded, proprioceptive, incentive, social etc."* [8].

Central to this thesis is the concept of *augmented feedback* (AFB), that is, information which would not be available otherwise, in this case provided by the computer. It is assumed that natural and augmented feedback operate on the same principles [84]. AFB can be *concurrent* (provided during performance) or *terminal* (provided after performance) [80][84].

Depending on the task and the form of AFB, it can either enhance, hinder or have no effect on learning [67][80]. For feedback to be effective, *"it must be both needed and understood"* [80] and it "*must be presented in a way that can be utilized by the learner to facilitate learning.*" [125]. A particular pitfall in feedback and training is that learners that constantly receive AFB may develop a dependency on the feedback [80]. According to Winstein *et al.*, *"The guidance hypothesis predicts that the guiding properties of augmented feedback are beneficial for motor learning when used to reduce error, but detrimental when relied upon.*" [135]. In this case, AFB may improve performance, but there may be zero or even negative transfer to situations where the feedback is not present [80][139]. Excessively abundant and specific feedback has also been criticised by Goodman *et al.* and Lange *et al.* [49][74].

In this thesis the focus is on motor skills that include both ballistic and feedback-controlled movements. Many rapid movements, such as striking an object, do not necessarily rely on corrective feedback. Once initiated, they proceed in a "ballistic" manner, without further adjustment. Longer and slower movements are often affected by feedback, but when practiced repeatedly, they become increasingly ballistic. However, even movements that have become ballistic can still be modified, disrupted, or stopped by feedback, but typically only after the first 200 ms [79].

Exertion interfaces and digital sports

More than half of this thesis, publications [P1],[P4],[P5], and [P6] deal with athletic skills and video mirror interfaces. The term "video mirror" refers to a setup like the one in Figure 2, where the camera view is shown on a screen in front of the user. Mirrors and video analysis are used in many sports to spot errors in pose and motion. In research literature, several computer-assisted motion and biomechanics analysis systems have been described. Various approaches include user-assisted video analysis, tracking devices, and computer vision [3][61][90][138]. There also exists commercial software for analyzing sport videos [113][134]. More examples of information technology for providing feedback for sports can be found, e.g., in the review by Liebermann *et al.* [75].

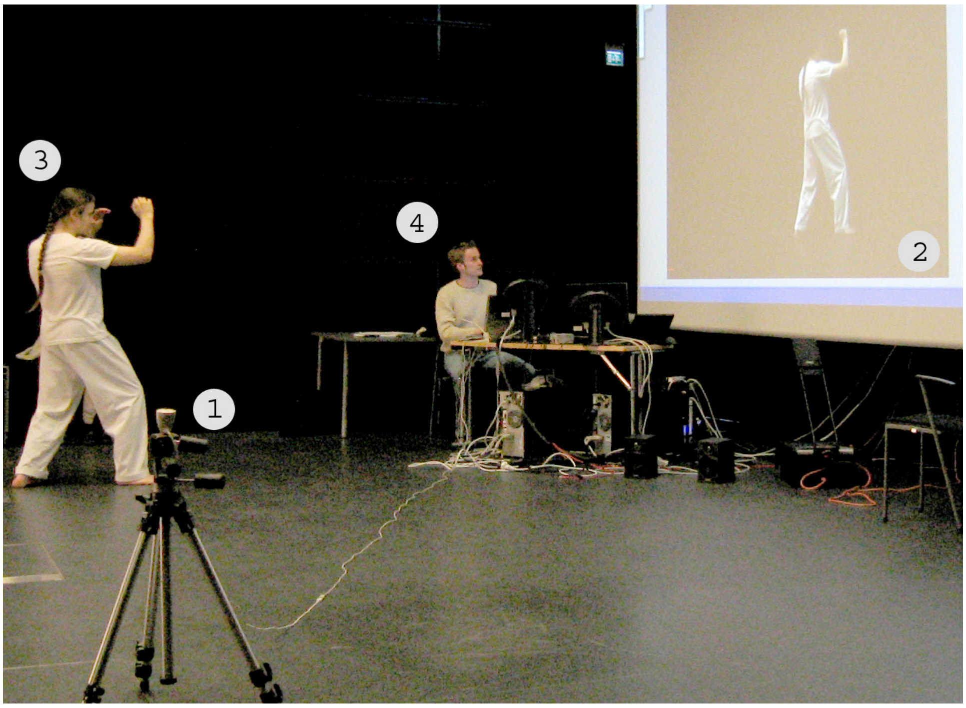

Figure 2. A test setup for an interactive video mirror for sports training, consisting of a camera (1), camera view projected on a screen (2), user (3) and test instructor (4).

Video mirrors can be seen as a light-weight, unencumbered alternative for virtual reality (VR) motor learning, which is a recent and active area of research. As technology has developed, feedback studies have moved on from verbal to computer generated and videotape feedback (VTFB). Menickelli's study indicates that *"both self- and instructor-controlled VTFB was more effective than either self- or instructor-controlled verbal KP in learning a multiple degrees-of-freedom skill."* [80] Todorov *et al.* showed that augmented feedback in VR can accelerate motor learning [126]. Todorov *et al.* have been followed by the Just Follow Me VR system by Yang *et al.*, where a 'ghost' model emerges from the learner's body [139], the Tai Chi training system by Chua *et al.* [23], and a study of the effect of AFB in VR on the drilling skills of dental students [131]. Non-VR motor training systems using computers and video projection include the anticipatory cues for tennis by Shim *et al.* [107], and the multimodal dance training system by Nakamura *et al.* [83], which consists of a combination of wearable vibro-tactile timing cue devices and a robot that moves a display in the dance space, indicating the desired amount of translation.

Training with the help of video mirrors or replays is a particular form of observational motor learning. According to Ferrari, "Observational learning of a motor skill involves: (1) observation of a model, which allows one to imitate and understand a modelled demonstration, and (2) self-observation, which allows one to actively regulate one's own learning and performance" [39]. It is not clear how observational learning works prior to actual movement, but "there is little doubt that a considerable amount of learning, particularly that which occurs early in practice, comes from studying and imitating the actions of others" [104]. Observation of a model can promote error detection and correction capabilities, and observers can detect timing patterns [104]. In relation to self-observation, Pollock suggests that observing a trainee instead of an expert can still facilitate learning [93].

In general, there has been much research in perceptive user interfaces and speech-based interaction. Interactive video mirrors and processed video of the user have been used in games and art installations starting from VideoPlace by Krueger et al. [72], where the two-dimensional video image of the user interacted with computer generated animated characters. Krueger calls the approach artificial reality, but it can also be considered as augmented reality, especially if the background of the user is not removed and graphics are only overlaid in the camera view. There are commercial applications of the approach, such as the Eye-Toy camera and related games for the PlayStation 2 games console [111]. The MIT Media Lab Alive system is an example of a more sophisticated 3D interactive video mirror where one can interact with computer generated characters using gestures [137].

There are also many previous examples of human-computer interfaces that require physical exertion. In the wake of Konami's Dance Dance Revolution video game released in 1998 [68], arcades all over the world feature dancing games where the player has to perform combinations of steps in rhythm to the music. Ishii *et al*. presented an "athletic-tangible interface", a ping-pong table that augments the sport with dynamic graphics and sound [63]. Mueller *et al*. discuss social aspects of networked computerized sports, based on a game setup with a regular soccer ball and life-size videoconference screen [82]. Chi *et al*. present a wearable sensor system that registers impacts in taekwondo competitions [22]. Physical interaction has also been a part of many dance and theatre performances, such as the various projects using EyesWeb [17][37], and the multi-camera, multi-projection mixed dance reality by Meador *et al.* [78].

2.2 Interactive video mirrors for sports training

Publication [P4] shows how interactive video mirrors can combine the benefits of mirrors and video in repeated performance and evaluation of acrobatic and martial arts moves. The video mirror setup is shown in Figure 2. The novelty is that motion analysis is integrated into the training with a full-body user interface so that there is no need to leave the training area to operate a computer or video equipment. Compared to VR motion training, the mirror setup provides a wireless and less encumbered interface.

If the camera is placed by your side, you can see yourself in a side view while looking forward, which helps in martial arts training where you should focus on the direction of your opponent. Using gesture and/or speech control, playback and recording functionality can be used without the athlete having to leave the training area to operate a computer or video equipment. This allows the user to focus on the learning task.

Compared to videotapes and real mirrors, video mirrors have the advantage that the delay of feedback can be freely adjusted. Visual monitoring and movement reproduction delays play a complex role in motor learning and performance. In a study by Carrol and Bandura, observed action patterns were reproduced significantly better with concurrent visual monitoring than monitoring delayed by 75 seconds [18]. Concurrent monitoring, e.g., with help of a mirror, can help in immediate spotting and correcting of errors. Watching yourself in a mirror can also give you courage in moves where you would not otherwise know your distance from the ground, e.g., when doing a Capoeira style cartwheel [1]. On the other hand, it is difficult to correct ballistic moves on the fly. When practicing kicks with a mirror, one first kicks, then figures out what went wrong and tries again. For spinning kicks where you can not keep your eyes on the screen, it can also be beneficial to add a few seconds of delay so that you can perform the kick and then immediately see it played back. Skill retention can also be improved by delayed reproduction of a model performance, which means that you have to retain, retrieve and produce an action instead of simply imitating the model concurrently [104].

Publication [P4] compares the following three interfaces for controlling the recording and playback:

- A system that estimates the amount of motion made by the user. If the motion exceeds a threshold, the system starts recording. When there is no motion for a specific time interval, recording stops and the system begins to play the recorded video in a loop, showing every other replay in slow-motion.
- Speech commands. The available commands are ‘record’, ‘play’, ‘stop’ and ‘delay’, corresponding to the possible states of the system. In playback mode, the recorded video plays in a loop with every other replay in slow-motion, similar to the automatic system. In delay mode, live video from the camera is shown, delayed by a user-defined amount.
- Overlay buttons and sliders operated with gestures, as shown in Figure 3. The widgets (controls) move with the user, staying at a constant distance. The widgets are derived from the motion activated buttons overlaid in the camera view in Eye-Toy games [111]. The framing of the camera view is looser than in Eye-Toy to give the user room to perform large motions. Because of the framing, static buttons could not be reached from all locations.

In the user study with eight martial artists in [P4], speech control was ranked best four times and gesture control three times. The automatic system had problems when a move ended in a different position than it started from. The test subjects often stopped for a while and started watching the replay, but then moved back to starting position, which switched the system back to recording mode prematurely.

A clear advantage of speech commands is that commands can be given from any location and pose. Speech commands are also similar to how one normally interacts with a person using a video camera. However, the limitations of speech recognition technology must be considered in practice. Current speech recognition technology is robust only when training alone and not speaking to fellow trainers. Noise from the

[1] When learning a cartwheel, people often feel that they need to see how they place their hands on the floor. In Capoeira, however, you are often instructed to keep your eyes on your opponent.

environment and consumer grade hardware may also cause problems.

The overlay widgets appear to be a promising approach, although the target environment must be considered carefully. For example, one user criticized the overlay widgets for the reduced image size resulting from the split screen. This can present a problem, particularly if a monitor is used instead of a projected display. The widgets were also touched unintentionally during high kicks and cartwheels. It was also noticed that with a camera placed by the user's side, the widgets are at first reached for in screen coordinates instead of world coordinates, that is, when a widget is in front of you, you try to extend your hand sideways to reach it.

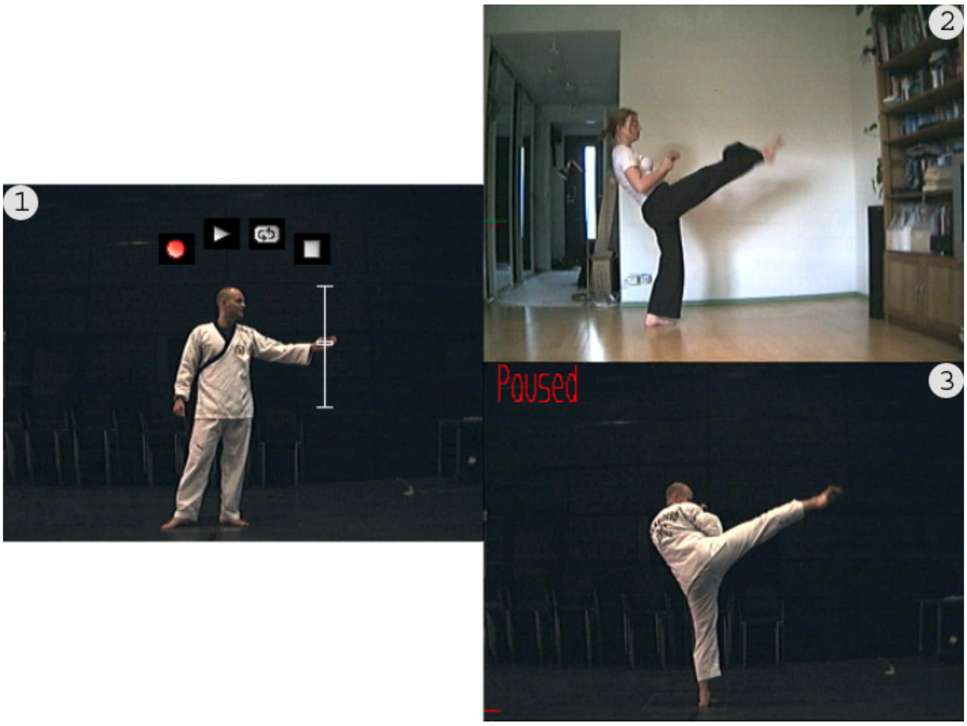

Figure 3. Inspecting individual frames of recorded video with a gesture operated slider that controls the playback position. The display is split into the camera view with overlay widgets (1), model view (2) and inspected frame (3)

2.3 Mixed reality martial arts: Kick Ass Kung-Fu

Publication [P5] continues the [P4] with the goal of motivating training through playful entertainment. The publication describes the design and evaluation of Kick Ass Kung-Fu, a martial arts game installation where one fights virtual enemies with kicks, punches, and acrobatic moves such as cartwheels. The game is shown in action in Figure 4, Figure 5, and Figure 6. With real-time image processing and computer vision, the video image of the user is embedded in 3D graphics on a virtual playfield with virtual opponents. The player's movements are exaggerated so that one can dodge the opponent's attacks by jumping several meters in the air and landing behind the opponent's back. Using dual projected screens, one at each end of the playing field, one can also continue by counter-attacking the enemy from behind.

Figure 4. Kick Ass Kung-Fu, a mixed reality martial arts game, at Kiasma Theater during PixelACHE 2004 (top) and 2005 (bottom)

Figure 5. The Kick Ass Kung-Fu computer vision is flexible. One can play together with a friend and use a variety of weapons.

Figure 6. Shouting supercharges the player in Kick Ass Kung-Fu.

Kick Ass Kung-Fu is a hybrid of games, sports, and performance. It has been exhibited at various media art festivals, e.g., at Kiasma Theatre during PixelACHE 2004 and 2005. The game is an example of designing for spectator experience instead of user experience only [97]. In a theatre show, we typically invite some skilled martial artists to do a show, but the audience can step on stage too. The game was also used in the Finnish television show Taistelevat Julkkikset, where celebrities fought against virtual, cartoonized celebrities.

Compared to previous work, the novelties of Kick Ass Kung-Fu are:

- Computer vision technology that accurately measures impacts but at the same time does not require wearable markers or sensors and allows practically any weapon to be used (a sword, *nunchucks*, thrown objects, etc.), and any number of simultaneous players who collaborate against the virtual opponents.
- One can shout to supercharge oneself with *chi* (*ki*) energy, as shown in Figure 6. Martial arts often emphasize the importance of *kiai* (shout) in conjunction with attacks. Shouting can intimidate the opponent, and makes the abdominal muscles flex, preparing the body for a counterattack.
- A novel video mirror setup with two screens and the camera by the player's side. This provides a good view of many kicks, and lets the audience see both the real and virtual versions of the game events. Combined with the robust computer vision, the setup also supports collaborative playing and training. It is easy to take turns and assess one's own performance while watching others play. In terms of observational learning, Kick Ass Kung-Fu facilitates both observing a model and self-observation.
- Exaggerated motion, meaning that one can jump higher and run faster on-screen. This can make one feel more skilled than in real-life, which can help motivate beginners.
- The game provides a unique opportunity to apply film-style martial arts techniques and acrobatic moves (cartwheels, backflips). Many people like to practice aesthetic martial arts tricks that are too risky to be used in real-life sparring. In Kick Ass Kung-Fu, the tricks can be put to use.
- Dynamic slow-motion. In a traditional game played with a gamepad, slow-motion is trivial to implement since one can press buttons to control the avatar even if it moves at different speeds. In a physically interactive system with a one-to-one mapping between the user and the avatar, slowing down the avatar makes it go out of sync with the user. Kick Ass Kung-Fu employs slow motion dynamically so that when you jump high enough, the camera feed is slowed down, and when you land, the feed is fast forwarded back to real time. One can first fight with normal speed on the ground, then perform a flamboyant slow-motion jump kick and continue with a series of punches that get shown at faster than normal speed.

The exaggerated motion and slow-motion were inspired by martial arts films and games. David Bordwell uses the term *expressive amplification of movement* when analyzing the aesthetic of Hong-Kong action movies [12]. *"Expressive amplification is one major way in which the Hong Kong film, as the fans like to say, goes over the top"* [12]. Bordwell links Hong-Kong movies to Eisenstein, who "*believed that every movement activates the entire body, and so in theatre and film one must 'sell' each action by exaggerating the body's role in forming it…This sort of stylized clarity,* *verging on cartoon movement, is quite different from the more subdued performance style characteristics of contemporary Hollywood acting, which tends to emphasize the face rather than the entire body."* [12]

Publication [P5] describes Kick Ass Kung-Fu and a user study with 46 martial artists. Publication [P6] examines Kick Ass Kung-Fu in the context of an open exhibition where people could come to play the game freely. [P5] is complemented with observations and interviews of children playing the game. Children were observed to have more difficulties in adapting to the profile view, but they were socially less inhibited, e.g., considering the shout interaction – we had to turn the shouting based supercharge off because of gallery workers' complaints about constant noise. For adults, the supercharging works as an icebreaker – people are shy at first, but when someone has the courage to shout and pose, the audience cheers and applauds.

Furthermore, publication [P6] discusses the combined effects of full-body interaction and violence in entertainment software. We propose that game age ratings should be modified to also consider the realism of the users' actions, instead of the current focus on audiovisual realism of harmful content, such as violence. The audiovisual focus can be explained by the history of rating guidelines for non-interactive media, such as movies. In interactive media, both user input and system output contribute to the user experience.

To minimize harm, it seems that the input and output should be consistently realistic or unrealistic. It is relatively harmless to play a game with unrealistic cartoon violence using a game pad, but as the realism of the user's actions increases, so does the importance of realistic consequences. It is generally considered harmful that video games do not create realistic consequences for the aggressor [46] and the violent behavior of the player is simultaneously rehearsed, rewarded and reinforced in various ways including the addition of extra points, sounds and game levels [33]. Embodied interaction seems to add a new pitfall regarding the consequences – when punching air, one does not learn to control the amount of force one puts behind the moves. This was demonstrated by a four-year-old boy who punched his father when telling him about Kick Ass Kung-Fu, although no harm was apparently intended.

2.4 Precursor: QuiQui's Giant Bounce

Publication [P1] has been included in this thesis as a precursor to the video mirrors and Kick Ass Kung-Fu. The publication presents QuiQui's Giant Bounce (Kukakumma Muumaassa), a game where the user controls a dragon that flies when the one waves one's hands and breathes fire when one shouts, as shown in Figure 7. The game works on a personal computer equipped with a webcam and a microphone.

Figure 7. In Kukakumma Muumaassa (QuiQui's Giant Bounce), the avatar flies when the user waves his or her hands and breathes fire when the user shouts.

Before QuiQui, Freeman *et al.* developed several game interfaces where the user controls an avatar with body movements [43][44][45]. To my knowledge, QuiQui was the first computer game to let one express oneself as the avatar using both body movements and voice (beyond speech commands). The goal was to get children moving when playing computer games and to support the development of their physical abilities and motor skills. This was a fairly novel concept at the time of the game's release as a free download (2001), but it has now become mainstream with the launch of the Sony Eye-Toy camera and games for the PlayStation 2 console.

Considering feedback and motor learning, QuiQui did not attempt to support practicing any specific movements. The game had a less elaborated goal of simply making children move. The user's actions were not clearly and unambiguously visualized on screen, as the cartoon avatar had only a discrete set of poses. Viewing the project in the light of the later work, this seems its greatest flaw. The visual feedback can be ambiguous, and the user has to explore the expressive capabilities of the avatar by trial and error. Considering the avatar as a model for observational learning, it could be thought that the users would start to mimic the avatar's movements. However, our experiences do not support the assumption, as shown in Figure 8. Although the game was usually found enjoyable, some users "got stuck" in poses that the avatar could not copy. Getting stuck can be explained by approaching motor learning as a multidimensional optimization problem, so that information (e.g., feedback), is used to "*channel the search through the perceptual-motor workspace to locate a task-relevant solution to the coordination function*" [84]. The avatar's response to user motions can be thought to provide gradient information about the error or fitness function of the optimization. A zero gradient makes the search difficult, which can happen if the user experiments with motions and poses so far from the avatar's repertoire that the avatar does not move at all.

Figure 8. Test subjects learning to fly in QuiQui's Giant Bounce. The figures at the bottom are blurred because of rapid movements. The flying avatar has a rigid body and three possible hand poses: up, horizontal and down, but this clearly does not prevent users from experimenting other poses and motions.

2.5 Computer games played by singing

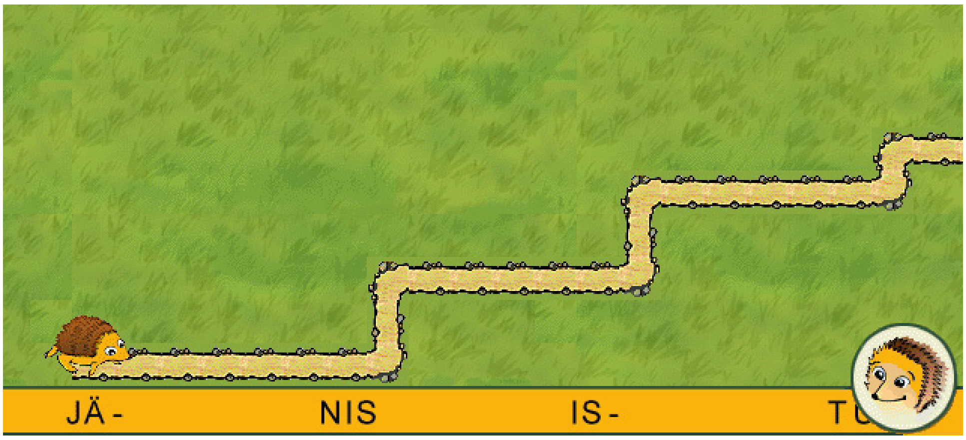

Figure 9. Screenshot of a pitch-controlled karaoke game. The vertical position of the hedgehog is controlled by the pitch of the user's voice. If the user sings the song correctly, the hedgehog stays on the path twisting along the melody. The words of the song are shown at the bottom of the screen.

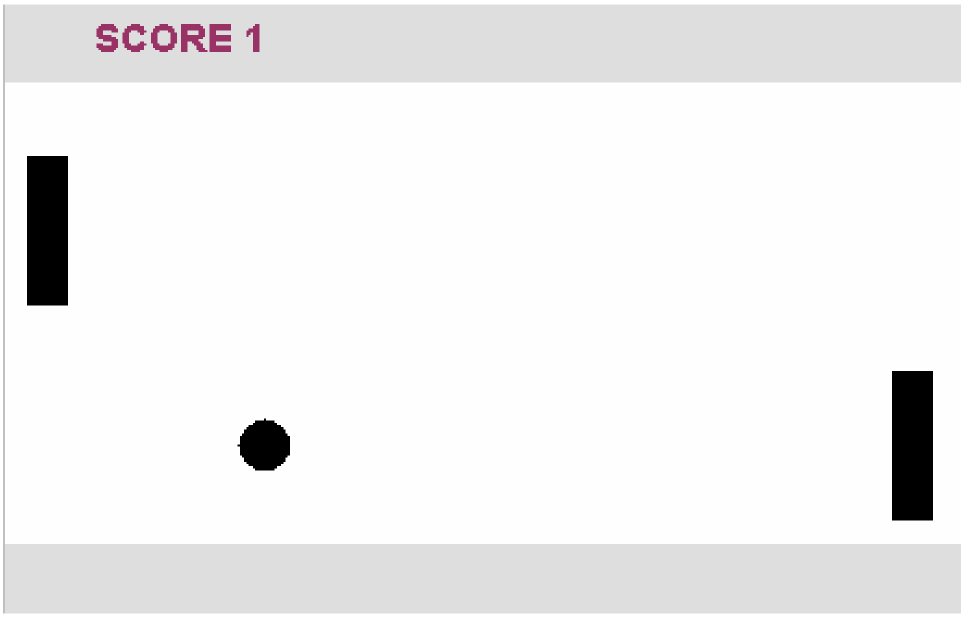

Figure 10. Screenshot of a pitch-controlled Pong-style game. The bats (black rectangles) move vertically according to the pitch of the user's voice. The screenshot is from one-player mode, where the bat on the right is controlled by the computer.

Publication [P3] describes two computer games played by singing into a microphone, shown in Figure 9 and Figure 10. The games were developed at Elmorex Ltd., where I worked as part of the PlaySingMusic team. PlaySingMusic was a musical edutainment (education and entertainment) game for children. Although the games were developed in 1999 and PlaySingMusic was released as Soittopeli in Finland in 2000, nothing could be published until 2004, when rival products had appeared.

The games provide visual feedback that can help in learning to sing in tune and control one's voice. Instead of trying to distinguish pitch, one can see it because it gets mapped onto properties of visual objects in the game graphics. Pitch has been defined as the characteristic of a sound that makes it sound high or low or determines its position on a scale [101].

Human perception of pitch is a complex process, but in the case of singing, pitch corresponds to the frequency at which the vocal cords vibrate [129]. The concept of pitch and related adjectives, such as “high” and “low” are not always clear for people with no musical training. For example, we found that when trying to sing “higher”, people may only change the vowel they are voicing, for example, from ‘a’ to ‘e’, which can be explained in terms of spectral changes that contribute to the perception of pitch [101].

In the Hedgehog game in Figure 9, pitch controls the vertical position of the hedgehog running on the horizontally scrolling background. The player’s goal is to keep the hedgehog on the path that twists according to the melody of the song playing in the background. Here, the ideas of Welch *et al.* are applied in the context of a narrative and a game. Welch *et al.* were the first to propose real-time visualization of pitch trajectories for singing education, and show that visual AFB can help in learning to sing in tune [55][130]. The hedgehog and the path are metaphors for the actual and desired pitches. They provide concurrent AFB, and the prize awarded at the end of the song provides terminal AFB. Publication [P3] also contributes in the form of discussing the problems of pitch detection technology when applied to controlling an avatar (see Section 3.2).

Howard and Welch suggest that exploring pitch contours is beneficial for the development of singing [55]. The Pong game in Figure 10 provides this kind of exercise. The game features visual feedback similar to the Hedgehog game, but there is no desired melody to be sung. Pitch is mapped to the vertical position of the bat, and the player tries to intercept the bouncing ball with it.

The Pong game provides an example of combining turn-based and continuous real-time control so that two players can share a microphone to play against each other. In two-player mode, pitch controls both bats so that they are always at the same vertical position. The players take turns in singing or humming so that once a player hits the ball, it is the other player’s turn to react and sing. In multiplayer games, one has to use multiple control devices, or take turns using a shared control. Examples of the former are Tekken and other martial arts games played with several gamepads. Stewart et al. have also researched these kinds of interfaces for collaborative work [115]. Having two microphones would make it easy to implement a two-player game, because the pitch detection usually outputs the pitch of the loudest voice. However, few users have the equipment needed to connect two microphones to a soundcard.

2.6 Visual feedback for animation: Animaatiokone

Figure 11. Front and side views of Animaatiokone, a machine for creating clay animation. Animating takes place inside the transparent studio dome, where one can find a camera, set pieces, modelling clay, and buttons for capturing and deleting frames, and previewing the animation.

Publication [P2] describes Animaatiokone, an installation for creating clay animation, shown in Figure 11. The main design goal was to hide the complexities of technology and to make animating as easy as possible, so that novice users could experiment and learn about clay animation in a casual and fun environment, for example when waiting for a movie in the lobby of a movie theatre. The installation is based on a personal computer, a webcam, and a microphone.

Animation has become omnipresent in digital media, thanks to software tools for creating and distributing animation. An early interactive 2D computer animation system by Baecker [2] was soon followed by a scripted 3D system [20] and then a 3D animation system for people inexperienced with computers [50]. Animation tools have then evolved to the current software used to create web content, games, and movies. A 2.5D animation system has also been described [76]. However, although many 2D animation systems use scanned, hand-drawn pictures as the starting point [38], all research reported on 3D animation deals with computer generated animation and there are no studies of computerizing traditional clay or puppet animation.

Animaatiokone animations form a continuous story: The next animator can continue from where the previous one finished or start a new scene, but the previous animations cannot be deleted. The overhead display allows other people to watch the animator at work and users also often share each other's clay actors, created on a table beside the installation. The animations can be viewed at the installation's homepage, where one can check out what plot twists the next animators came up with and what perils one's actors ended up in.

In general, there are several reports of experimental audiovisual tools that explore new areas of expressiveness and facilitate social interaction and peer-to-peer communication. Animaatiokone can be seen as part of this branch of research, along with, e.g., I/O Brush and KidPad [54][102]. I/O Brush allows one to pick textures from the environment like colours from a palette, using a camera embedded in a brush. KidPad is a drawing/sketching tool with a shared zoomable canvas and multiple input devices. Clay animation is also highly tangible, in which sense Animaatiokone is related to tangible computing pioneered by Ishii *et al.* [62].

For the beginner, clay animation is easy to approach. Instead of manipulating curves and surfaces with a mouse, one can simply grab a handful of plasticine and start modelling. Compared to shooting animation on film or video, a computer can provide concurrent visual feedback, which helps in improving performance in the same way as in the singing games and Kick Ass Kung-Fu.

Animaatiokone provides low-latency terminal AFB through previews that do not require rendering. For concurrent AFB, Animaatiokone has an automatic onion-skinning display that was evolved through iterative prototyping and testing with end users. The low-tech prototype is shown in Figure 12. In the onion-skinning window, the camera view and the previous frame are blended together, which helps in realizing how much the clay figures move between frames. At the first frame of each animation, the onion-skinning view shows only the camera view. Publication [P3] also presents a novel, more powerful onion skinning control for blending up to three images. The control was left out of the installation to simplify the interface, but it is used in the recently released Animaatiokone desktop software.

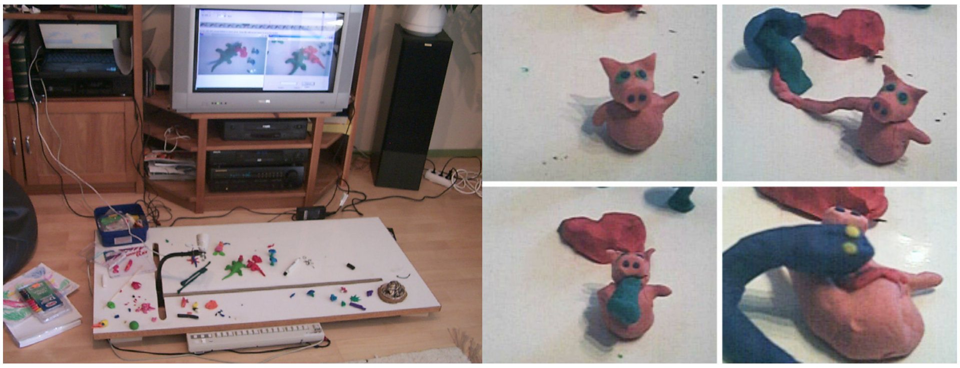

Figure 12. Left: The low-tech prototype of Animaatiokone consisted of a television, a whiteboard with a camera rail and a simplified PC keyboard. Right: Sample frames of animation created with the prototype.

3 TECHNOLOGY

Human-computer interaction can be defined as dialogue, a cyclic process where both the user and the computer express themselves and process information [25]. Technology sets practical limits to the expressiveness of the participants of the conversation. The limits are usually harder on the input side of the user interface. The computer speaks but it does not listen - audiovisual rendering technology has matured to a stage where consumer PCs can output fairly realistic 3D audio and graphics, but we can not talk and gesture to the computer as naturally as in human to human communication. This thesis explores and in some cases expands the expressive freedom enabled by computer vision and audio analysis technology.

3.1 Background

Input and output devices

Cameras and microphones are still fairly novel input devices for HCI, despite some thirty years of research and development. The first camera-based interactive system I know of was a digital musical instrument called DIMI-O, developed as early as 1971 by Finnish inventor Erkki Kurenniemi [73]. DIMI-O featured musical synthesis controlled by the video signal of a video camera. Another early example is the camera based motion tracker system VNS (Very Nervous System) by media installation artist David Rokeby [29]. Rokeby created his first motion tracking systems in 1983 and has been improving the technology ever since, moving from dedicated hardware to software implementations (softVNS). In the HCI literature, a better known pioneer of the field is Myron Krueger with his seminal VideoPlace and other experiments [70][71][72].

For the singing games and Animaatiokone, microphones and cameras are the obvious input devices, since they are used in non-computerized singing and clay animation too. For QuiQui and Kick Ass Kung-Fu, other input devices could be used, e.g., magnetic or gyroscopic position and rotation sensors, and wearable flex sensors. Computer vision was selected because it enables a natural and unencumbered user interface that can be robust and accessible outside laboratory conditions. Wearable sensors and markers were ruled out, because they are awkward if several users want to play at the same time or by taking turns.

Regarding the computer's expressiveness, the thesis projects utilize conventional display devices and ordinary stereo speakers. No technological innovations are presented, and the research focuses more on what and how to display. Visualization and interaction using large screens, as in the case of Kick Ass Kung-Fu, is an active area of research [66][117]. Although 3d displays and immersive sound systems (see, e.g., Cruz-Neira *et al.* [27] and Lokki *et al.* [77]) would have provided interesting design options, conventional hardware was chosen to create systems that could be used with reasonable cost and effort outside laboratory conditions, e.g., in homes, galleries, and at movie theatres.

Practical computer vision

Computer vision and hearing are ideal technologies for the interaction designer, since they mimic the primary senses we use in communication. In many application contexts, they enable the game to sense all that the interaction designer senses. In practice, machine perception is still far behind human perception. Practical computer vision systems may partly model human perception, but they are also engineered to utilize prior knowledge available about the application and the problem domain. Computer vision is both an art and a science – there are many ad-hoc systems without sound statistical or mathematical basis [30].

Although computer vision is a relatively new field of research, there are already several comprehensive textbooks and surveys [1][30][81][110]. There are also books on specific topics, such as motion analysis by fitting curves to image contours [9]. Many methods start with common preliminary processing operations, such as edge-finding [110] or background subtraction [127]. Background subtraction means that the image is segmented into foreground and background pixels, often based on a statistical model of the background. Other common building blocks are digital signal processing methods, optimization, principal component analysis (PCA), Bayesian methods, and least-squares theory, including the widely used Kalman filter. In addition to computer vision textbooks, the fundamentals are covered in textbooks on statistics, signal processing and numerical methods [48][53][95].

In the context of this thesis, the relevant computer vision topic is human motion capture. In addition to user interfaces, human motion capture is useful for surveillance, motion capture for films and motion analysis for rehabilitation and sports, to name a few. The surveys by Moeslund [81] and Aggarwal and Cai [1] are good starting points for getting familiar with the literature.

Motion capture methods can be divided into two categories:

- Model-based computer vision. A model, such as a deformable contour template or a skeleton is fitted to perceived image features, e.g., the silhouette of the user.
- Motion capture by contextually assigning meaning to low-level image features, as in the game interfaces by Freeman *et al.* [43][44][45]. For example, a computer game can be controlled with a coloured glove by first finding the pixels with colour close to the glove, and then mapping the location of the mass centre of the pixels to the position or velocity of an avatar. This kind of vision technology is heavily based on simplifying assumptions about the context of use, in this case that the glove is the only visible object with the colour of interest.

Model fitting methods search for a combination of model parameters that maximizes a fitness or probability function or minimizes an error function. Crucial to real-time model fitting is the dimensionality of the parameter space. For example, Wilhelms *et al.* use a model of 70 body segments, with a total of 75 degrees of freedom [132]. The search space has several local optima, which makes the approach impractical for real-time applications. In Wilhelms' case, a human assists the fitting, which is practical in non-real-time applications, such as motion capture for movie post-production.

For real-time model fitting, it is usually assumed that the tracked person or persons perform only a subset of motions so that the search can be confined into a practically small portion of the parameter space. The assumptions can be implemented as constraints for optimization, or the dimensionality of the model can be reduced using statistical methods, such as PCA [9][110]. Furthermore, the search can be constrained by assuming that the motion is reasonably continuous, so that the model parameters for the next image frame can be predicted from the previous ones, e.g., with a Kalman filter [9][110].

This thesis uses low-level contextual vision, because from the point of view of usability, the degrees of freedom and the expressiveness of the user must not be limited, but real-time operation in real-life conditions must still be achieved. For example, in QuiQui's Giant Bounce, we observed a wide variety of unanticipated movements, shown in Figure 8. A model that could explain all possible movements was considered too complex for real-time model fitting.

In addition to model-based and low level approaches, computer vision human motion capture can be divided into markerless/marker-based, real-time/offline, single/multiple camera, and user-assisted/autonomous methods. Current systems achieve good results, given enough computing time and/or multiple cameras [109][123]. Unconstrained single-camera real-time markerless 3D motion capture on consumer computers remains an unsolved problem. However, this may change in the near future, thanks to advances in multimodal parameter estimation and stochastic optimization methods, such as particle filters [7][36] and importance sampling [56][109]. These methods generate random model parameter vector samples that converge on the peaks of a fitness function or a parameter probability density function. For each sample, the model must be rasterized and the results have to be compared to the camera input to evaluate the fitness or probability. This can be computationally intensive, but the massively increasing computing power of GPU:s (graphics processors) provides interesting possibilities for efficient implementation of the methods [87].

Practical computer hearing: voice as sound paradigm

Perhaps the most common audio analysis task for HCI is speech recognition. However, in this thesis, low-level audio features, specifically signal power and pitch are used as continuous control signals. In 2001, at the same time as QuiQui's Giant Bounce was released as a free download, Igarashi and Hughes introduced this approach to the HCI community as "voice as sound" [60]. One of their example applications is a television volume control where the user says "Volume up, ahhh" and the volume continues to increase as long as the "ahhh" continues. Earlier, similar features of sound have mainly been used to control musical user interfaces. For example, in the Swim Swan performance in 1993, the pitch and loudness of a clarinet were used to control musical synthesis [47].

Speech input without actual speech recognition has been used in toys, such as a parrot that repeats words [106], for which it is necessary only to segment the input signal into words or phrases. The voice as sound approach is analogous to the contextual low-level computer vision discussed above.

3.2 Technological design of the thesis projects

Considering the technology of the thesis projects, Kick Ass Kung-Fu is based on a novel combination of basic computer vision methods. In QuiQui's Giant Bounce and in the singing games, the novelties are more related to finding a user interface approach that satisfies the requirements of the application within the constraints of the technology.

The user-avatar and avatar-virtual interface layers

Usability engineering for computer games is a new and active field of research [65]. The basic paradox is that a good user interface for utility software tries to make things easier for the user, but a game usually presents a challenge. According to many authors, challenge is one of the central elements of a game [92].

For user-centered design of embodied games such as QuiQui's Giant Bounce and Kick Ass Kung-Fu, it is useful to divide the interface into human-avatar and avatar-virtual interfaces [P1]. The human-avatar interface denotes the mapping of the user's actions to the avatar's actions. The avatar-virtual interface incorporates the interaction happening in the game world, for example, puzzles solved by manipulating virtual objects.

In [P1] it is proposed that for a game to be usable, the human-avatar interface should be as intuitive and transparent as possible. The human-avatar interface should allow the user focus on the challenge presented by the avatar-virtual interface, to allow the user think and act from the point of view of the avatar. This is related to Gold's observation that in the context of games, an interface does not simply denote a system of widgets and control devices between the user and the program. Gold uses a doll as an example:*"The doll/child relationship is different: the child does not interface with the exterior of the doll to reach some deeper 'dollness', and I believe the same holds for video games.*" [89]. It is a usability issue if the user, as his or her real-world self, does not know how to move so that the avatar moves to a certain direction. However, it is a more general game design problem, for example, to design an opponent that presents an optimal challenge in a martial arts game.

It may seem that designing an intuitive human-avatar interface is trivial – one simply needs to make the avatar copy the user's movements. However, even if technology was perfect, this would only be possible if the real and the virtual environments matched one-to-one. When the user is standing and the avatar is flying, as in the case of QuiQui's Giant Bounce, the player-avatar mapping and suitable technology is best found through iterative prototyping and testing with end users.

User-centered design of real-time computer vision for children: QuiQui's Giant Bounce

To choose the right computer vision method one needs to carefully identify and specify the requirements for the technology [30]. For HCI, *"Usability determines the requirements for technological innovation"* [26]. Since technology is not yet perfect, the task is to find a set of simplifying assumptions that allow a computationally efficient computer vision solution, but do not sacrifice usability.

Crowley et al. specify three main requirements for human-centred perceptual user interfaces: robustness and autonomy, low latency, and privacy protection [26]. The following additional requirements were defined for QuiQui's Giant Bounce [P1]:

1. Completely automatic operation without any learning stages, initialization or settings that need user participation. Our goal is that children can use the game without adult guidance.
2. The methods must adapt rapidly to changes in the environment, including lighting and camera position.
3. The methods must adapt to the differences in various camera models, including frame rates in the range 15...30 fps, noise, motion blur and colour resolution.
4. The system must tolerate several visible humans, either one player and viewers or several users participating collaboratively.
5. The user must not be required to wear any specific clothing or markers.
6. Low computational complexity to enable maximal computational resources for the actual game application.

Before QuiQui, there was little written about designing perceptual computer games for children in uncontrolled real-life environments, such as homes and daycare centres. The previous systems were either installations for galleries etc. [11] or were not tested extensively with children [43][44][45]. The Intel & Mattel games were targeted for homes [32], but they required a static background, which violates requirement 2.

We developed an initial vision-based interface prototype and tested it using a novel usability evaluation method called *peer tutoring,* described in detail by Höysniemi *et al.* [59]. The basic idea was that in order to confirm that children can play the game without adult supervision, we observed children teaching each other to play. The prototype was found non-intuitive, and we redesigned it to cope with the most frequent way children tried to fly in the test: They waved their hands to fly and controlled the direction of the flight by bending their bodies. Later on, we have also used Wizard of Oz (WOz) prototyping in which a human observes the motions of the user and controls the game with a keyboard [57]. Video recorded at the WOz sessions can be used for testing purposes when the actual computer vision system is developed.

The computer vision solution for QuiQui uses image moments (statistical moments of the pixel distribution of an image) as the salient features. The solution is a variation of moment-based vision investigated by Freeman *et al.* [43][44][45]. To make the system adapt rapidly to changes in the environment, background subtraction is not used. Instead, the analysis is based on the differences of pixel intensities between consecutive frames, as shown in Figure 13. The angle of the user's body is determined by a line drawn through a left-biased and right-biased mass-center of the nonzero pixels in the thresholded difference image. The thrust for the flying is proportional to the number of nonzero pixels.

Figure 13. The computer vision in QuiQui. Left: original input. Center: absolute value of the difference of pixel intensities in this and the previous video frame. Right: thresholded difference and the line corresponding to the moments.

Combining background subtraction and pyramidal optical flow: Kick Ass Kung-Fu

In QuiQui's Giant Bounce, a cartoon avatar mimics the user's movements. This is possible in real-time because of the simplified motion of the avatar, for which only the angle of the user's body and the amount of motion are required as the control signals. In Kick Ass Kung-Fu, the player must be able to perform as wide a variety of martial arts and acrobatic techniques as possible, which demands a different approach. Similar to Krueger's VideoPlace [72] and the games by Intel and Mattel [32], the video figure of the user is embedded inside computer graphics after background subtraction, as shown in Figure 4. This makes the human-avatar interface transparent – the user *is* the avatar. The approach also provides accurate visual feedback of the user's movements.

There is not much written about computer vision in applications where the user's video image directly interacts with computer graphics, probably because in many cases the technology involved is trivial, especially in a controlled environment. It is often enough to test whether the user's silhouette and a visual object overlap, e.g., to make a virtual soap bubble burst. For more elaborated interaction, a model can be fitted to the silhouette. However, in an unconstrained case, the mapping between the silhouette and the model is inherently many-to-many [15]. Pfinder uses a model of coloured Gaussian blobs in combination with background subtraction, apparently with good results, but it assumes that there is only one user [136].

In Kick Ass Kung-Fu, the velocities of the outline pixels of the user silhouette are found by computing pyramidal Lucas-Kanade optical flow [86], as shown in Figure 14. All shapes that move fast enough generate impulses to objects that they collide with. The impulses are proportional to the magnitude and direction of the optical flow. Contrary to model-based methods that could be applied in real-time, this low-level contextual vision system allows any number of simultaneous users with practically no limitations to the movements or weapons that can be used. No initialization phase is necessary between games or users taking turns.

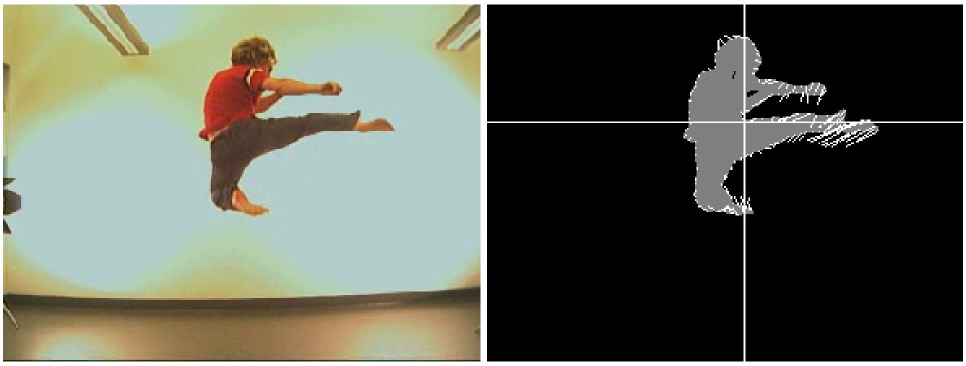

Figure 14. Camera view (left) and computer vision debug view of a jump kick in Kick Ass Kung-Fu. The crosshair shows the mass center of the user pixels and the lines show the optical flow computed.

Computer hearing: voice pitch and intensity as the control signals

In QuiQui's Giant Bounce and Kick Ass Kung-Fu, the signal power of the microphone input is used to trigger actions: QuiQui breathes fire when the user shouts [P1] and in Kick Ass Kung-Fu, the players can supercharge themselves by shouting [P6]. Technologically, this is trivial, as long as the application is designed so that false triggers have no negative consequences for the user. False triggers can be caused by background noise, other people than the user, or sound generated by the application itself.

In the singing games described in this thesis, the pitch of the user's voice controls the vertical position of an avatar [P3]. Pitch detection (fundamental frequency estimation, f0 estimation) has been researched widely since the 70's [31][96][120]. With current computers, many pitch detection methods can be implemented in real-time. The actual algorithms used in PlaySingMusic cannot be disclosed. Publication [P3] discusses pitch detection on a general level, and focuses on aspects of the user interface that are visible when playing the games.

For the game or interaction designer it is important to note that no pitch detection method is perfect. The performance is affected by differences in people's voices, the quality of the microphones and soundcards, and interfering sounds from the environment. Publication [P3] observes that a majority of the errors are octave errors and proposes a solution: Pitch is transposed inside one octave, or to the octave closest to the predicted/desired pitch. The publication also discusses the pros and cons of the approach – for example, octave transposing ensures that people with differently ranged voices can play together. On the other hand, it breaks the one-to-one mapping between the pitch and the visuals, which causes ambiguous visual feedback. For example, if the user makes his or her voice quickly jump exactly an octave, the avatar does not move at all. This was puzzling for at least one user who tested the pong game.

4 DISCUSSION

The aim of this thesis was to explore what kinds of novel educational software could be created with audiovisual signal processing technology applied to user interfaces. As a result, a series of user interface prototypes were developed and user studies were conducted. The ideas for the prototypes were the result of the simultaneous observation of the following:

- The needs or problems of a specific user group, e.g., the difficulty of being aware of one's body in martial arts training. Awareness is increased by practicing in front of a mirror, but it is difficult to look at the mirror during certain moves, e.g., spinning kicks.
- The potential of signal processing and computer vision technology in solving the problems and satisfying the needs, e.g., by constructing a "video mirror" that consists of a projected screen, a video camera, and a computer [P4]. The computer can delay the video feed so that one can observe a spin kick immediately after completion. The video replays can also be controlled with gestures without leaving the training area to operate a computer or video equipment.

There are practically limitless possibilities beyond the cases presented in this thesis. The cases were selected based on personal values (education), interests (sports and audiovisual arts), and the resources available. Despite the limited scope, the following tries to answer the research questions by inspecting the common properties of the cases.

4.1 Psychomotor edutainment and interface technology

As a whole, the thesis projects demonstrate how novel interfaces and technology can widen the scope of *psychomotor edutainment*, a form of edutainment relatively neglected so far. I use the term psychomotor as a reference to one of the three domains of learning in the taxonomy of educational objectives by Bloom *et al.* [10][69]:

- Cognitive (intellectual abilities and skills, recall or recognition of knowledge)
- Affective (interests, attitudes, values)
- Psychomotor (motor skills)

Bloom *et al.* divide educational objectives into classes or levels of increasing complexity. For example, the classes for the cognitive domain are knowledge, comprehension, application, analysis, synthesis, and evaluation. The taxonomy was created to help teachers and other people who deal with curricular and evaluation problems.

Bloom *et al.* did not complete the analysis of the psychomotor domain. Harrow presents one possible division [52]:

1. Reflex
2. Fundamental movements: applicable mostly to young children (crawl, run, jump, reach, change direction)
3. Perceptual abilities: catch, write, balance, distinguish, manipulate
4. Physical abilities: stop, increase, move quickly, change, react
5. Skilled movements: play, hit, swim, dive, use
6. Non-discursive communication: express, create, mime, design, interpret

Considering traditional computer games and edutainment, the psychomotor domain is largely unexplored, with the exception of creative software for music and visual arts. Action video games can also be considered to train hand-eye-coordination, at least at the lowest levels of Harrow's taxonomy.

The projects described in this thesis increase the diversity of skills that can be practiced in a digital environment. QuiQui's Giant Bounce and the singing games are mostly related to Harrow's levels 1 – 5. The games are goal-oriented, and leave little room for improvisation and creative expression. On the other hand, Kick Ass Kung-Fu and Animaatiokone cover at least some aspects of all the levels. In its current form, Kick Ass Kung-Fu allows the player to use a variety of fighting styles. Instead of trying to get the highest possible score, people often prefer to experiment with moves and watch themselves on the screens.

From the point of view of HCI research it can be noted that the limiting factor in the use of computers for psychomotor edutainment is the user interface. Pivec *et al.* suggest that learning activities should be mapped to interface actions, and learning concepts should be mapped to interface objects [92]. Using novel interfaces and interface technology, the thesis work expands the scope of psychomotor edutainment for arts and sports. In effect, the user interface *is* the educational content, in agreement with David Rokeby's notion of interface as content [29]. Skelly has also noted that *"video games are virtually all interface."* [89].

4.2 Action realism: the convergence of simulators and edutainment

Traditionally, psychomotor learning has had more emphasis in simulators than edutainment. In a way, the thesis work represents a convergence of the two. For a unified view of edutainment and simulators, a central concept seems to be *realism*.

Combining education and fun is generally not easy, but even if a game is designed for entertainment, one can learn from realistic game elements. When one learns to play a game, realism has an effect on which aspects of the gained skills and knowledge are applicable outside the game. Realistic war games can educate about weaponry and tactics. In Wildlife Rescue, *"players who take the time to learn about their animals by using the animal manuals are given useful information about food, medical conditions, treats, and general animal facts."* [100]. In the context of simulators, *fidelity* is a more common term than realism. According to

Williams, *"In simulations where skilled operation is a desired outcome, high levels of fidelity will result in the user being able to more readily transfer learned procedures to the real working environment."* [133].

On the other hand, reduced fidelity, detail and complexity can sometimes be more desirable for beginners [16][94]. Oblinger talks about a trade-off between realism and fun as *"Games tend to be less real but more fun; simulations have greater realism."* [85]. A major attraction of computer games is that one can try out things not possible in the real world. The same applies to action movies, which would be quite dull if there were no unrealistic stunts.

In conjunction with Kick Ass Kung-Fu we have used the terms *user interface realism* and *action realism* [P5][P6]. Kick Ass Kung-Fu is an example of psychomotor edutainment based on action realism: a game too unrealistic to be considered a simulation, but rather edutainment seeking the optimal balance between realism and fantasy, or learning and fun. The game employs the principle of selective realism, that is, only the elements necessary to the goals of the learning exercise are provided [64][85]. The user interface of Kick Ass Kung-Fu is realistic in that players actually perform martial arts moves and get visual feedback of their performance. The motivating fantasy elements are the exaggerated motion and supercharge effects.

4.3 Critique: transfer of learning

The transfer of learning can be questioned in all the applications and prototypes developed for this thesis in light of the guidance hypothesis – the users can become dependent on the concurrent augmented feedback. However, user response has been highly positive, and people have shown strong interest in training their skills using the tools. The heart rates measured show that Kick Ass Kung-Fu can provide intensive exercise [P5], so that there can be health benefits even if the game is not optimal for learning real-life martial arts. Sedighian and Klawe suggest that to improve learning, feedback should be removed gradually [105], but it remains a future research goal to find a way to achieve this, e.g., in Kick Ass Kung-Fu. The player could possibly fight blindfolded, relying on sounds and intrinsic feedback.

On the other hand, as the boundaries between simulations, edutainment and games are blurring, so is the distinction between simulated and real environments. There may be no need for a skill to be applied in an environment where augmented feedback is not present, since computers and games are ubiquitous in our everyday life. Many people enjoy trying to get a high-score in a singing game at home together with friends, but have no interest in performing publicly or otherwise applying the obtained skill in "real" life.

The Staraoke game and television format is an example of the fusion of training and performing with augmented concurrent visual feedback. People compete in a singing game on television, so that it is natural to use the PC version of the game for practicing at home [114]. Similarly, when Kick Ass Kung-Fu is exhibited on stage in a theatre, it can be regarded as a digital sport of its own, rather than a learning environment for traditional sports. For sports that share things in common, such as TaeKwonDo and Karate, practicing one may support the practicing of another. It seems logical to conclude that the same can apply to a combination of a digital and a traditional martial art.

4.4 Design guidelines and lessons learned

I will now venture to conclude this discussion with some design guidelines and lessons learned. One should note that the conclusions are not based on hard scientific data. At best, they are observations and hypotheses to be validated through future work.

Considering computer vision, I learned the hard way that one should aim only as high as the application requires. It will still take some years for computers to become powerful enough for a fully general-purpose single-camera human tracking system. With limited computing power, the more information one tries to extract from the video signal, the less reliable the information tends to become. For example, evaluating the fit of a complex model on an image is computationally intensive, which can limit the frame rate that can be achieved. Simplifying the evaluation, e.g., by using lower resolution or fewer optimization iterations typically leads to noisy tracking results. The results can be filtered to reduce noise, but this reduces temporal resolution, analogous to a lower frame rate. It is best to simplify the technology based on application-specific assumptions. However, validating the assumptions and the solution typically requires careful testing with end users.

I have also learned that one should not program a computer to do a man's work unless it is necessary. Wizard of Oz –style interaction works not only in usability testing, but also in end applications. In particular, people tend to have a better understanding of drama than computers. In Kick Ass Kung-Fu, it was easiest to first implement keyboard controls for the enemies instead of designing artificial intelligence. It turned out that people often want to use the keyboard to play against their friends. Even though we eventually programmed the AI, we still use the keyboard in television appearances and other occasions where timing and character are important. Furthermore, it was a lot easier to design the AI after concretely playing the role of the enemies for hours. Kick Ass Kung-Fu also provides automatic slow-motion replays of final blows, but the keyboard can be used to manually trigger a replay when something interesting happens.

This thesis describes digitally enhanced experiences based on real-life activities such as singing or kung-fu. There is no obvious way to select the starting points, but at least one should probably avoid designing a digitalized experience inferior to the original one. I think that Animaatiokone, Kick Ass Kung-Fu and the singing games succeed in that they preserve the essence of the original experience, and only enhance it with feedback, scoring and special effects. Kick Ass Kung-Fu does not deprive the player of the joy of physical activity. Instead, it reduces risk, allowing the player to apply impractical but aesthetic moves in combat. As a counterexample, I would think twice before attempting to design an embodied snowboarding game. Even if an appropriate control scheme could be developed, it would be difficult to compensate for the loss of experiencing real snow and sun, unless of course the game was designed for off-season practicing.

4.5 Future directions

In the future I would like to further investigate the social aspects of embodied games and digital sports. The audience or a group of peers is what really seems to make a game like Kick Ass Kung-Fu work. Stepping in front of an audience may take courage but it can be empowering as well. Discussion and commenting promote innovation in playing techniques and styles. In addition to developing commenting and communication tools, such as uploading of game videos to a website, I would like to study more closely how the game is perceived. For example, it would be interesting to use an eye-tracking system to measure how the audience's attention is divided between the screen and the players. This could inform the design of the game setup and visuals. Regarding the visuals, I would like to explore the aesthetic and cognitive effects of cinematic techniques such as overlapping editing, borrowing ideas from the expressive amplification of movement of Hong Kong action movies [12].

On the technical side, I am still working on computer vision based human motion capture. I would like to develop a tracker suitable for recognizing martial arts and gymnastics techniques in Kick Ass Kung-Fu. This would enable at least elementary style-based scoring, which would support creative expression and exploring of moves, making the game more interesting for the spectators and the player community. Currently, the more enemies you defeat, the more points you get. It would also be very interesting to try out physical enhancements, such as ropes with which you could swing and backflip, or a power track – a long trampoline that could span the whole playing field.

5 MAIN RESULTS AND MY CONTRIBUTION

The main results of the thesis and my contribution can be summarized as follows. None of the publications [P1-P4] have previously formed a part of another thesis. Publications [P5] and [P6] are also included in the thesis of Johanna Höysniemi. All the publications have been peer-reviewed based on publication-ready manuscripts. The reviews were single-stage, except for [P5], published in the ACM CHI 2005 conference, where the revisions made after the first round were reviewed. The ACM CHI papers [P2] and [P5] have probably the biggest impact factors. According to Citeseer analysis of the impact of publication venues in computer science, ACM CHI proceedings are the most influential HCI publication, surpassing journals such as ACM TOCHI and Interacting with Computers [24].

Publication [P1]: A Computer Vision and Hearing Based User Interface for a Computer Game for Children

The paper describes a novel embodied computer game where one can express oneself as the avatar using both gestures and voice. The paper also defines a set of requirements for user-centered computer vision technology, elaborating on previous discussion. The computer vision system described in the paper is rather simple and its use is limited beyond the interface in question, but it serves as an example of a technical solution that satisfies the specified requirements. I designed the user interface in collaboration with Johanna Höysniemi, designed and implemented the computer vision and hearing technology, programmed the game and wrote all of the text in the paper.

Publication [P2]: Animaatiokone - an Installation for Creating Clay Animation

The paper describes a computerized clay animation system. This contributes to the considerable amount of research on software tools for creative audiovisual expression. The paper discusses the final system and user-centered design process, including design principles, technology, prototyping, testing, and lessons learned. I designed the user interface in collaboration with Ari Nykänen and Mikko Lindholm, conducted the usability studies, designed and implemented the software and electronics, and wrote all of the text in the paper.

Publication [P3]: Musical Computer Games Played by Singing

The paper discusses action computer games where the player sings into a microphone and the pitch of the player's voice is analyzed in real-time and used to control the game. The paper presents observations and lessons learned from developing and testing two games, and explains the requirements and restrictions of pitch detection for such games. I designed the pitch controlled pong game and wrote all of the text in the paper, except for paragraphs 2-3 of Section 4.1.

Publication [P4]: Interactive Video Mirrors for Sports training

The publication investigates how video "magic mirror" interfaces can be used for sports training. The paper compares different interfaces: speech commands, buttons and sliders operated with gestures, and an automatic system that records all movement and plays the recorded video when there's no movement. The interfaces provide novel visual feedback that speeds up repeated performing and analyzing of sports techniques, such as spin kicks and cartwheels. I designed and implemented the interfaces, conducted the user study and wrote all of the text.

Publication [P5]: Martial Arts in Artificial Reality

Here, we continue [P4] by introducing Kick Ass Kung-Fu, a martial arts game based on a video mirror interface. The novelties of the interface are that there are two projected screens, and the motion of the player is exaggerated. We also describe how slow-motion can be used in a video mirror interface. Furthermore, we present a novel combination of computer vision techniques that allows any number of simultaneous users and sets practically no limits to their movements or weapons used. I designed and implemented the user interface and technology, conducted the user study together with Johanna Höysniemi and Tommi Ilmonen, and wrote all of the text.

Publication [P6]: Children's and Parents' Perception of Full-Body Interaction and Violence in a Martial Arts Game

The paper continues [P4] and [P5] by studying Kick Ass Kung-Fu in the context of an open exhibition where people could come play the game freely. [P5] described a study of Kick Ass Kung-Fu played by martial artists, and here we complement the study with observations and interviews of children playing the game. The paper also contributes by discussing the combined effects of full-body interaction and violence. I developed the new features in the game, conducted the study with Johanna Höysniemi and wrote a third of the text.

BIBLIOGRAPHY

[1] Aggarwal, J.K. and Cai, Q. 1999. Human Motion Analysis: a Review. *Computer Vision and Image Understanding: CVIU*, Vol. 73, No. 3, 428-440, available electronically at http://citeseer.ist.psu.edu/aggarwal99human.html, link checked 1st March 2006

[2] Baecker, R. 1980. Picture Driven Animation. In Freeman, H. (ed.) *Tutorial and selected readings in Interactive Computer Graphics*, IEEE Computer Society, 273-288. Originally published in *Proceedings of Spring Joint Computer Conference, AFIPS* (1969)

[3] Baecker, R., Miller, D., Reeves, W. 1981. Towards a Laboratory Instrument for Motion Analysis. *Computer Graphics*, vol. 15, No. 3, August 1981, 191-197

[4] Balzer, W.K, Doherty, M.E., O'Connor, R. 1989. Effects of Cognitive Feedback on Performance, *Psychological Bulletin,* Vol. 106, No. 3., 410-433

[5] Bederson, B. B. 2000. Fisheye menus. *Proceedings of the 13th Annual ACM Symposium on User interface Software and Technology (UIST '00)*, San Diego, California, United States, November 06 - 08. ACM Press, New York, NY, 217-225

[6] Bederson, B. B., Meyer, J., and Good, L. 2000. Jazz: an extensible zoomable user interface graphics toolkit in Java. *Proceedings of the 13th Annual ACM Symposium on User interface Software and Technology (UIST '00)*, San Diego, California, United States, November 06 - 08. ACM Press, New York, NY, 171-180.

[7] Bernier, O., Cheung-Mon-Chan, P. 2006. Real-Time 3D Articulated Pose Tracking using Particle Filtering and Belief Propagation on Factor Graphs, *Proceedings of BMVC'06*. Available at http://www.bmva.ac.uk/bmvc/2006/papers/ 159.pdf, link checked 18th Feb 2007

[8] Bilodeau, E.A. and Bilodeau, I.M. 1961. Motor Skills Learning, *Annual Review of Psychology,* Vol. 12, 243-280

[9] Blake, A. and Isard, M. 1998. Active Contours: The Application of Techniques from Graphics, Vision, Control Theory and Statistics to Visual Tracking of Shapes in Motion, Springer-Verlag

[10] Bloom, B.S. (ed.), Engelhard, M.D., Hill, W.H., Furst, E.J., and Krathwohl, D.R. 1956. *Taxonomy of educational objectives. The Classification of Educational Goals. Handbook I: Cognitive Domain*, David McKay Company

[11] Bobick, A. F., Intille, S. S., Davis, J. W., Baird, F., Pinhanez, C. S., Campbell, L. W., Ivanov, Y. A., Schütte, A., and Wilson, A. 2000. Perceptual user interfaces: the KidsRoom. *Communications of the ACM*, Vol. 43, No. 3 (Mar. 2000), 60-61.

[12] Bordwell, D. 2001. Aesthetics in Action: *Kungfu*, Gunplay, and Cinematic Expressivity, in Yau, C.M. (ed.), *At Full Speed Hong Kong Cinema in a Borderless World,* University of Minnesota Press, 73-94

[13] Bracken, P., and Thomas, P. 2002. Time to move beyond the mind-body split, *BMJ*, Vol. 325, 1433-1434

[14] Bradski, G.R. 1998. Computer Vision Face Tracking For Use in a Perceptual User Interface, *Intel Technology Journal* Q2, 1998.

[15] Brand, M. 1999. Shadow Puppetry. *Proceedings of International Conference on Computer Vision*, 1237-1244

[16] Caird, J.K. 1996. Persistent issues in the application of virtual environment systems to training. *Proceedings of Third Annual Symposium on Human Interaction with Complex Systems (HICS '96)*, 124-132

[17] Camurri, A., Coletta, P., Massari, A., Mazzarino, B., Peri, M., Ricchetti, M., Ricci, A., Volpe, G. 2004. Toward real-time multimodal processing: EyesWeb 4.0, *Proceedings of AISB 2004 Convention: Motion, Emotion and Cognition*, Leeds, UK, March 2004.

[18] Carrol, W.R., and Bandura, A. 1985. Role of Timing of Visual Monitoring and Motor Rehearsal in Observational Learning of Action Patterns. *Journal of Motor Behavior,* Vol 17, No. 3, 269-281

[19] Cartesian dualism in Wikipedia, *http://en.wikipedia.org/wiki/Cartesian_dualism,* link checked 30th March 2006

[20] Catmull, E. 1972. A system for computer generated movies, *Proceedings of the ACM annual conference*, Vol. 1, ACM Press

[21] Chang, B., and Ungar, D. 1993. Animation: from cartoons to the user interface. *Proceedings of UIST '93*, ACM Press, New York, NY, 45-55.

[22] Chi, E.H., Song, J., Corbin, G. 2004. "Killer App" of wearable computing: wireless force sensing body protectors for martial arts, *Proceedings of UIST 2004*, ACM Press, 277-285

[23] Philo Tan Chua, Crivella, R., Daly, B., Ning Hu, Schaaf, R., Ventura, D., Camill, T., Hodgins, J., Pausch, R. 2003. Training for Physical Tasks in Virtual Environments: Tai Chi, *Proceedings of IEEE Virtual Reality 2003*, 87-94

[24] Citeseer estimated impact of publication venues in computer science, *http://citeseer.ist.psu.edu/impact.html*, link checked 31st March 2006

[25] Crawford, C., 2003, *The Art of Interactive Design*, No Starch Press

[26] Crowley, J. L., Coutaz, J., and Bérard, F. 2000. Perceptual user interfaces: things that see. *Communications of the ACM* 43, 3 (Mar. 2000), 54-64.

[27] Cruz-Neira, C., Sandin, D.J., DeFanti, T.A. 1993. Surround-screen projection-based virtual reality: the design and implementation of the CAVE, *Proceedings of SIGGRAPH'93*, ACM Press, 135-142

[28] Damasio, A. 2001. *Descartesin virhe* (Finnish version, originally *DESCARTES' ERROR. Emotion, Reason and the Human Brain*, 1994), Terra Cognita, Finland

[29] David Rokeby home page, *http://homepage.mac.com/davidrokeby/home.html*, link checked 12.12.2005.

[30] Davies, E. R. 2005. Machine vision : theory, algorithms, practicalities - 3rd ed. Elsevier

[31] De Cheveigné, A., and Kawahara, H. 2002. YIN, a fundamental frequency estimator for speech and music. *Journal of the Acoustic Society of America*, 111 (4)

[32] D'Hooge, H. 2001. Game Design Principles for the Intel Play Me2Cam* Virtual Game System. *Intel Technology Journal* Q4, 2001.
[33] Dill, K.E., and Dill, J.C. 1998. Video game violence: review of the empirical-literature. *Aggression and Violent Behavior*, 3 (4), 407-428.
[34] Dourish, P. 2001. Where the Action Is - The Foundations of Embodied Interaction, MIT Press
[35] Druin, A. 1999. Cooperative inquiry: developing new technologies for children with children. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: the CHI Is the Limit (CHI '99)*. Pittsburgh, Pennsylvania, United States, May 15 - 20, 1999. ACM Press, New York, NY, 592-599.
[36] Duetscher, J., Blake, A., and Reid, I. 2000. Articulated Body Motion Capture by Annealed Particle Filtering. *Proceedings of 2000 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'00)*, Vol. 2, p. 21-26
[37] EyesWeb, http://www.eyesweb.org, link checked 22nd Feb 2006
[38] Fekete, J-D., Bizouarn, É., Cournarie, É., Galas, T., Taillefer, F. 1995. TicTacToon: A Paperless System for Professional Animation. *Proceedings of 22nd annual conference on Computer Graphics and Interactive Techniques (SIGGRAPH'95)*, ACM Press, 79-90
[39] Ferrari, M. 1996. Observing the Observer: Self-Regulation in the Observational Learning of Motor Skills. *Developmental Review,* Vol. 16, 203–240
[40] Finney, S.A., 1997. Auditory Feedback and Musical Keyboard Performance. *Music Perception*, vol. 15, no. 2, pp 153-174, 1997
[41] Fogelholm, M. 2005. Lihavat lapset saatava liikkeelle. *Helsingin Sanomat,* 6th March 2005, Sanoma Osakeyhtiö, Finland
[42] Freed, A., Chaudhary, A., Davila, B. 1997. Operating Systems Latency Measurement and Analysis for Sound Synthesis and Processing Applications. *Proceedings of International Computer Music Conference (ICMC'97)*, 1997
[43] Freeman, W.T., Anderson, D., Beardsley, P., Dodge, C., Kage, H., Kyuma, K., Miyake, Y., Roth, M., Tanaka, K., Weissman, C., Yerazunis, W. 1998. Computer Vision for Interactive Computer Graphics. *IEEE Computer Graphics and Applications*, May-June, 1998, pp. 42-53
[44] Freeman, W. T., Beardsley, P. A., Kage, H., Tanaka, K., Kyuma, K., Weissman, C. D. 1999. Computer vision for computer interaction. *SIGGRAPH Comput. Graph.* 33, 4 (Nov. 1999), 65-68.
[45] Freeman, W.T., Tanaka, K., Ohta, J., Kyuma, K. 1996. Computer vision for computer games. *Proceedings of IEEE 2nd Intl. Conf. on Automatic Face and Gesture Recognition,* Killington, VT, October 1996
[46] Funk, J.B. and Buchman, D.D. 1995. Video game controversies. *Pediatric Annals*, Vol. 24, 91-94.
[47] Furukawa, K. and Dutilleux, P. 2002. Live-electronics Algorithms in the Multimedia Work 'Swim Swan'. *Proceedings of the 5th International Conference on Digital Audio Effects (DAFx-02)*

[48] Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B. 2004. *Bayesian Data Analysis*, 2nd edition, Chapman & Hall / CRC

[49] Goodman, J.S., Wood, R.E., Hendrickx, M. 2004. Feedback Specificity, Exploration, and Learning. *Journal of Applied Psychology,* Vol. 89, No. 2, 248–262

[50] Hackathorn, R.J. 1977. Anima II: A 3-D Color Animation System. *Computer Graphics*, Vol. 11, No. 2

[51] Hanna, L., Risden, K., Alexander, K. 1997. Guidelines for usability testing with children. *interactions* 4, 5 (Sep. 1997), 9-14.

[52] Harrow, A.J. 1972. *A Taxonomy of the Psychomotor Domain.* David McKay Company

[53] Hayes, M.S. 1996. Statistical Digital Signal Processing and Modeling. John Wiley & Sons

[54] Hourcade, J. P., Bederson, B. B., Druin, A., Taxén, G. 2002. KidPad: collaborative storytelling for children. *CHI '02 Extended Abstracts on Human Factors in Computing Systems*. ACM Press, 500-501

[55] Howard, D.M., and Welch, G.F. 1989. Visual feedback applied to the learning of conscious pitch control in singing. *Proceedings of Eurospeech 89, European Conference on Speech Communication and Technology*, Paris, France, 26-28 Sept (1989), 485-488.

[56] Hämäläinen, P., Aila, T., Takala, T., Alander, J. 2006. Mutated kd-tree importance sampling for optimization. *Proceedings of the Ninth Scandinavian Conference on Artificial Intelligence (SCAI 2006)*.

[57] Höysniemi, J., Hämäläinen, P., Turkki, L., and Rouvi, T. 2005. Children's intuitive gestures in vision-based action games. *Communications of the ACM,* 48, 1 (Jan. 2005), 44-50.

[58] Höysniemi, J., Hämäläinen, P. and Turkki, L. 2004. Wizard of Oz Prototyping of Computer Vision Based Action Games for Children. *Proceedings of International Conference on Interaction Design and Children* (IDC2004), 1-3 July 2004, Maryland, USA, 27-34.

[59] Höysniemi, J., Hämäläinen, P., Turkki, L. 2003. Using peer tutoring in evaluating the usability of a physically interactive computer game with children. *Interacting with Computers*, Volume 15, Issue 2, April 2003.

[60] Igarashi, T., and Hughes, J. F. 2001. Voice as sound: using non-verbal voice input for interactive control. *Proceedings of the 14th Annual ACM Symposium on User interface Software and Technology (UIST '01)*, Orlando, Florida, November 11 - 14. ACM Press, New York, NY, 155-156.

[61] Iredale, F., Farrington, T., Jaques, M. 1997. Global, fine and hidden sports data: applications of 3-D vision analysis and a specialised data glove for an athlete biomechanical analysis system. *Proceedings of Mechatronics and Machine Vision in Practice*, 260-264

[62] Ishii, H., and Ullmer, B. 1997. Tangible bits: towards seamless interfaces between people, bits and atoms. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '97). ACM Press, 234-241.

[63] Ishii, H., Wisneski, C., Orbanes, J., Chun, B., Paradiso, J. 1999. PingPongPlus: design of an athletic-tangible interface for computer-

supported cooperative play. *Proceedings of CHI'99*, ACM Press, 394-401

[64] Jackson, R.L., and Fagan, E. 2000. Collaboration and learning within immersive virtual reality. *Proceedings of the Third international Conference on Collaborative Virtual Environments*. E. Churchill and M. Reddy, Eds. CVE '00. ACM Press, 83-92.

[65] Jørgensen, A. H. 2004. Marrying HCI/Usability and computer games: a preliminary look. *Proceedings of the Third Nordic Conference on Human-Computer interaction (NordiCHI '04)*, ACM Press, 393-396

[66] Khan, A., Matejka, J., Fitzmaurice, G., Kurtenbach, G. 2005. Spotlight: directing users' attention on large displays. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '05)*, ACM Press, 791-798.

[67] Kluger, A.N., and DeNisi, A. 1998. Feedback Interventions: Toward the Understanding of a Double-Edged Sword. *Current Directions in Psychological Science*, Vol. 7, Issue 3, 67-72

[68] Konami Dance Dance Revolution. http://www.konami.co.jp/en/ir/financialinfo/annualreport/report1999/am01.html, link checked 22nd Feb 2006

[69] Krathwohl, D.R., Bloom, B.S., Masia, B.B. 1964. Taxonomy of Educational Objectives. The Classification of Educational Goals. Handbook II: Affective Domain. David McKay Company

[70] Krueger, M. 1983. *Artificial Reality*. Addison-Wesley.

[71] Krueger, M. 1991. *Artificial Reality II*. Addison-Wesley.

[72] Krueger, M., Gionfriddo, T., Hinrichsen, K. 1985. VIDEOPLACE - An Artificial Reality. *Proceedings of CHI'85*, ACM Press, 35-40

[73] Kuljuntausta, P. 2002. *ON/OFF Eetteriäänistä sähkömusiikkiin*. Like kustannus ja Kiasma

[74] Lange, C., Unnithan, V., Larkam, E., Latta, P.M. 2000. Maximizing the benefits of Pilates-inspired exercise for learning functional motor skills. *Journal of Bodywork and Movement Therapies,* 4(2), 99-108

[75] Liebermann, D.G., Katz, L., Hughes, M.D., Bartlett R.M., McClements, J., Franks, I.M. 2002. Advances in the application of information technology to sport performance. *Journal of Sports Sciences,* Vol. 20, No. 10, 755-769

[76] Litwinowicz, P.C. 1991. Inkwell: A 2½-D Animation System. *ACM Computer Graphics*, Vol 25, No 4, 113-122

[77] Lokki, T., Savioja, L., Väänänen, R., Huopaniemi, J., Takala, T. 2002. Creating Interactive Virtual Auditory Environments. *IEEE Computer Graphics and Applications,* Vol. 22, No. 4, 49-57

[78] Meador, W.S., Rogers, T.J., O'Neal, K. Kurt, E., Cunningham, C. 2004. Mixing Dance Realities: Collaborative Development of Live-Motion Capture In a Performing Arts Environment. *ACM Computers in Entertainment*, Vol. 2, No. 2

[79] Mechner, F. 1995. *Learning and Practicing Skilled Performance*. Mechner Foundation, 1995.

[80] Menickelli, J. 2004. The Effectiveness of Videotape Feedback in Sport: Examining Cognitions in a Self-Controlled Learning

Environment. Ph.D. thesis. Available at http://etd.lsu.edu/docs/available/etd-04052004-160403/, link checked 21st Feb 2006
[81] Moeslund, Thomas B. 1999. *Computer vision-based human motion capture – a survey*. Aalborg University, Laboratory of Computer Vision and Media Technology, 1999. (LIA Report; LIA 99-02.-ISSN 0906-6233)
[82] Mueller, F., Agamanolis, S., Picard, R. 2003. Exertion interfaces: sports over a distance for social bonding and fun. *Proceedings of CHI'03*, ACM Press, 561-568
[83] Nakamura, A., Tabata, S., Ueda, T., Kiyofuji, S., Kuno, Y. 2005. Multimodal presentation method for a dance training system. *CHI '05 Extended Abstracts on Human Factors in Computing Systems*. ACM Press, New York, NY, 1685-1688
[84] Newell, K.M. 1991. Motor Skill Acquisition. *Annual Review of Psychology,* Vol. 42, 213-37
[85] Oblinger, D.G. 2004. The Next Generation of Educational Engagement. *Journal of Interactive Media in Education*, 2004 (8)
[86] OpenCV computer vision library, *http://sourceforge.net/projects/opencvlibrary/*, link checked 1st Mar 2006
[87] Owens, J.D., Luebke, D., Govindaraju, N., Harris, M., Krüger, J., Lefohn, A.E., Purcell, T.J. 2007. A Survey of General-Purpose Computation on Graphics Hardware. *Computer Graphics Forum,* Vol. 26
[88] Paluska SA, Schwenk TL. 2000. Physical activity and mental health: current concepts. *Sports Medicine,* Mar 2000, 29(3), 167-80.
[89] Pausch, R., Gold, R., Skelly, T., Thiel, D. 1994. What HCI designers can learn from video game designers. *Conference Companion on Human Factors in Computing Systems (CHI '94)*. ACM Press, 177-178
[90] Perš, J., Bon, M., Kovačič, S., Šibila, M., Deľman, B. 2002. Observation and Analysis of Large-scale Human Motion. *Human Movement Science*, 21(2), 295-311
[91] Pinhanez, C.S., Wilson, A.D., Davis, J.W., Bobick, A.F., Intille, S., Blumberg, B., Johnson, M.P. 2000. Physically Interactive Story Environments. *IBM Systems Journal*, Vol. 39, No 3-4.
[92] Pivec, M., Dziabenko, O., Schinnerl, I. 2003. Aspects of Game-Based Learning. *I-KNOW 03, the Third International Conference on Knowledge Management*, available electronically at http://www.unigame.net/flash/content/I-Know_GBL-2704.pdf, link checked 1st Mar 2006.
[93] Pollock, B.J., and Lee, T.D. 1992. Effects of the Model's Skill Level on Observational Motor Learning. *Research Quarterly for Exercise and Sport,* Vol. 63, No. 1, 25-29
[94] Prensky, M. 2001. *Digital Game Based Learning,* McGraw-Hill
[95] Press, W.H., Teukolsky, S.A., Vetterling, V.T., Flannery, B.P. 2002. *Numerical Recipes in C++*, 2nd ed. Cambridge University Press
[96] Rabiner, L.R., Cheng, M.J., Rosenberg, A.E. McGonegal, C.A. 1976. A Comparative Performance Study of Several Pitch Detection Algorithms. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. 24, No. 5

[97] Reeves, S., Benford, S., O'Malley, C., and Fraser, M. 2005. Designing the spectator experience. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '05)*, 741-750.
[98] Robinson, T.N. 1999. Reducing Children's Television Viewing to Prevent Obesity a Randomized Controlled Trial. *Journal of American Medical Association,* Vol 282, No. 16.
[99] Robinson, D., Groves, J. 2004. *Introducing Philosophy*, Icon Books
[100] Romero, C. 2003. Wildlife rescue as a framework for learning. *ACM Computers in Entertainment* 1, 1 (Oct. 2003)
[101] Rossing, T.D. 1990. *Science of Sound*, 2nd ed., Addison-Wesley
[102] Ryokai, K., Marti, S., Ishii, H. 2004. I/O brush: drawing with everyday objects as ink. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '04)*. ACM Press, 303-310.
[103] Sawchuk, A. A., Chew, E., Zimmermann, R., Papadopoulos, C., Kyriakakis, C. 2003. From remote media immersion to Distributed Immersive Performance. *Proceedings of the 2003 ACM SIGMM Workshop on Experiential Telepresence (ETP '03)*, Berkeley, California. ACM Press, New York, NY, 110-120.
[104] Schmidt, R.A., and Wrisberg, C.A. 2004. *Motor Learning and Performance*. Human Kinetics
[105] Sedighian, K., and Klawe, M.M. 1996. An interface strategy for promoting reflective cognition in children. *Conference Companion on Human Factors in Computing Systems: Common Ground (CHI '96).* ACM Press, 177-178.
[106] Selker, T. 1996. New paradigms for using computers. *Communications of the ACM*, Vol. 39, No. 8, 60-69
[107] Shim, J., Chow, J.W., Carlton, L.G., Chae, W-S. 2005. The Use of Anticipatory Visual Cues by Highly Skilled Tennis Players. *Journal of Motor Behavior*, Vol. 37, No. 2, 164–175
[108] Sinha, A.K., Landay, J.A. 2001. Visually Prototyping Perceptual User Interfaces through Multi-modal Storyboarding. *Proceedings of Workshop on Perceptual User Interfaces*, Nov. 15-16, 2001, Orlando, Florida
[109] Sminchisescu, C. 2002. Estimation Algorithms For Ambiguous Visual Models: Three-Dimensional Human Modeling and Motion Reconstruction in Monocular Video Sequences. PhD thesis, Institute National Politechnique de Grenoble (INRIA)
[110] Sonka, M., Hlavac, V., Boyle, R. 1999. *Image Processing, Analysis and Machine Vision*, 2nd edition. Brooks/Cole Publishing Company
[111] Sony Eye-Toy, *http://www.eyetoy.com*, link checked 9th Sep 2004
[112] Spacarino, F., Davenport, G., Pentland, A. 2000. Media in performance: Interactive spaces for dance, theater, circus, and museum exhibits. *IBM Systems Journal*, Vol 39, Nos 3&4, 479-510
[113] Sports Motion Inc., *http://www.sportsmotion.com*, link checked 21st February 2006
[114] Staraoke home page, *http://www.staraoke.fi*, link checked 2nd March 2006

[115] Stewart, J., Bederson, B.B., Druin, A. 1999. Single Display Groupware: A Model for Co-present Collaboration. *Proceedings of CHI'99*. ACM Press

[116] Subrahmanyam, K., Kraut, R.E., Greenfield, P.M., Gross, E.F. 2000. The Impact of Home Computer Use on Children's Activities and Development. *Children and computer technology*, Vol. 10, No. 2, Fall/Winter 2000

[117] Swaminathan, K., and Sato, S. 1997. Interaction design for large displays. *interactions* 4, 1 (Jan. 1997), 15-24

[118] Sweetser, P., and Wyeth, P. 2005. GameFlow: a model for evaluating player enjoyment in games. *Computers in Entertainment,* 3, 3 (Jul. 2005)

[119] Sykes, J., and Brown, S. 2003. Affective gaming: measuring emotion through the gamepad. *CHI '03 Extended Abstracts on Human Factors in Computing Systems*. ACM Press, New York, NY, 732-733.

[120] Tabrikian, J., Dubnov, S., Dickalov, Y. 2004. Maximum A-Posteriori Probability Pitch Tracking in Noisy Environments Using Harmonic Model. *IEEE Transactions on Speech and Audio Processing*, Vol. 12, No.1

[121] Tammisalo, O. 2005. Rakkauden evoluutio. Ihmisen parinvalinnan biologiaa. Terra Cognita, Finland

[122] Taylor CB, Sallis JF, Needle R. 1985. The relation of physical activity and exercise to mental health. *Public Health Reports*, Mar-Apr 1985, 100(2), 195-202.

[123] Theobalt, C., de Aguiar, E., Magnor, M. A., Theisel, H., Seidel, H. 2004. Marker-free kinematic skeleton estimation from sequences of volume data. *Proceedings of the ACM Symposium on Virtual Reality Software and Technology (VRST '04)*. ACM Press, 57-64

[124] Thomas, B.H., and Calder, P. 2001. Applying cartoon animation techniques to graphical user interfaces. *ACM Trans. Comput.-Hum. Interact.* 8, 3 (Sep. 2001), 198-222

[125] Thorpe, D.E., Valvano, J. 2002. The Effects of Knowledge of Performance and Cognitive Strategies on Motor Skill Learning in Children with Cerebral Palsy. *Pediatric Physical Therapy*, 14(1), 2-15

[126] Todorov, E., Shadmehr, R., Bizzi E. 1997. Augmented Feedback Presented in a Virtual Environment Accelerates Learning of a Difficult Motor Task. *Journal of Motor Behavior*. 29(2),147-158

[127] Toyama, K., Krumm, J., Brumitt, B., Meyers, B. 1999. Wallflower: Principles and Practice of Background Maintenance. *Proceedings of International Conference on Computer Vision*

[128] Ware, C. 1996. Moving motion metaphors. Conference Companion on Human Factors in Computing Systems: Common Ground (CHI '96). ACM Press, New York, NY, 225-226

[129] Welch, G.F., Himonides, E., Howard, D.M., Brereton, J. 2004. VOXed: Technology as a meaningful teaching aid in the singing studio. *Proceedings of Conference on Interdisciplinary Musicology (CIM04)*

[130] Welch, G.F., Howard, D.M., Rush, C. 1989. Real-time visual feedback in the development of vocal pitch accuracy in singing. *Psychology of Music*, 17(2), 146-157

[131] Wierinck, E., Puttemans, V., Swinnen, S., van Steenberghe, D. 2005. Effect of augmented visual feedback from a virtual reality simulation system on manual dexterity training. *European Journal of Dental Education,* Vol. 9, 10-16

[132] Wilhelms, J., Van Gelder, A., Atkinson-Derman, L., Luo, A. 2000. Human Motion from Active Contours. *Proceedings of the IEEE Workshop on Human Motion,* available electronically at *http://www.cse.ucsc.edu/~alison/ Publications/*, link checked 1st Mar 2006

[133] Williams, V. 2003. Designing Simulations for Learning. *E-Journal of Instructional Science and Technology*, Vol. 6, No. 1

[134] WINanalyze motion analysis software, *http://www.winanalyze.com*, link checked 21st Feb 2006

[135] Winstein, C.J., Pohl, P.S., Lewthwaite, R. 1994. Effects of physical guidance and knowledge of results on motor learning: support for the guidance hypothesis. *Research Quarterly for Exercise and Sport,* 65(4), 316-23

[136] Wren, C.R., Azarbayejani, A., Darrel, T., Pentland, A. 1997. Pfinder: Real-Time Tracking of the Human Body. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 19, No. 7, July 1997

[137] Wren, C.R., Spacarino, F., Azarbayejani, A., Darrel, T., Davis, J., Starner, Kotani, A.,Chao, C., Hlavac, M., Russel, K., Bobick, A., Pentland, A. 1998. Perceptive Spaces for Performance and Entertainment (Revised). *ATR Workshop on Virtual Environments*, April 1998, Kyoto, Japan. Available at *http://citeseer.ist.psu.edu/wren97perceptive.html*, link checked 9th March 2007

[138] Yamamoto, M., Kondo, T., Yamagiwa, T., Yamanaka, K. 1998. Skill recognition. *Proceedings of Automatic Face and Gesture Recognition*, 604 – 609

[139] Yang, U., Kim, G.J. 2002. Implementation and Evaluation of "Just Follow Me": An Immersive, VR-Based, Motion-Training System. *Presence*, Vol. 11, No. 3, 304 –323