

Tools and Experiments in Multimodal Interaction

Tommi Ilmonen



TEKNILLINEN KORKEAKOULU
TEKNISKA HÖGSKOLAN
HELSINKI UNIVERSITY OF TECHNOLOGY
TECHNISCHE UNIVERSITÄT HELSINKI
UNIVERSITE DE TECHNOLOGIE D'HELSINKI

Tools and Experiments in Multimodal Interaction

Tommi Ilmonen

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 14th of December, 2006, 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 HUT

Tel. +358-9-451 2870

Fax. +358-9-451 5014

<http://www.tml.hut.fi/>

Available in electronic format at

<http://lib.tkk.fi/Diss/2006/isbn9512285517>

©Tommi Ilmonen

printed version:

ISBN-13 978-951-22-8550-1

ISBN-10 951-22-8550-9

ISSN 1456-7911

electronic version:

ISBN-13 978-951-22-8551-8

ISBN-10 951-22-8551-7

ISSN 1455-9722

Otamedia

Espoo 2006

ABSTRACT

Author Tommi Ilmonen
Title Tools and Experiments in Multimodal Interaction

The goal of this study is to explore different strategies for multimodal human-computer interaction. Where traditional human-computer interaction uses a few common user interface metaphors and devices, multimodal interaction seeks new application areas with novel interaction devices and metaphors. Exploration of these new areas involves creation of new application concepts and their implementation. In some cases the interaction mimics human-human interaction while in other cases the interaction model is only loosely tied to the physical world.

In the virtual orchestra concept a conductor can conduct a band of virtual musicians. Both the motion and sound of the musicians is synthesized with a computer. A critical task in this interaction is the analysis of the conductor motion and control of the sound synthesis. A system that performs these tasks is presented. The system is also capable of extracting emotional content from the conductor's motion. While the conductor follower system was originally developed using a commercial motion tracker, an alternative low-cost motion tracking system was also made. The new system used accelerometers with application-specific signal processing for motion capture.

One of the basic tasks of the conductor follower and other gesture-based interaction systems is to refine raw user input data into information that is easy to use in the application. For this purpose a new approach was developed: FLEXible User Input Design (FLUID). This is a toolkit that simplifies the management of novel interaction devices and offers general-purpose data conversion and analysis algorithms.

FLUID was used in a virtual reality drawing applications AnimaLand and Helma. Also new particle system models and a graphics distribution system were developed for these applications. The traditional particle systems were enhanced by adding moving force fields that interact with each other. The interacting force fields make the animations more lively and credible.

Graphics distribution become an issue if one wants to render 3D graphics with a cost-effective PC-cluster. A graphics distribution method based on network broadcast was created to minimize the amount of data traffic, thus increasing performance.

Many multimodal applications also need a sound synthesis and processing engine. To meet these needs the Mustajuuri toolkit was developed. Mustajuuri is a flexible and efficient sound signal processing framework with support for sound processing in virtual environments.

Keywords gestural interaction, conductor following, virtual reality, digital art, graphics clusters, particle systems, 3D sound, digital signal processing

TIIVISTELMÄ

Tekijä Tommi Ilmonen
Otsikko Multimodaaliset käyttöliittymät — menetelmiä ja kokeiluja

Tämän tutkimuksen tarkoitus on selvittää erilaisia menetelmiä multimodaaliseen/keholliseen ihmisen ja tietokoneen väliseen vuorovaikutukseen. Siinä missä perinteiset järjestelmät pohjautuvat tavallisimpiin laitteisiin (hiiri, näppäimistö) ja vuorovaikutusmenetelmiin (komentorivi, ikkunointi), multimodaalisten käyttöliittymien sovellusmahdollisuuksia etsitään uusien laitteiden ja menetelmien kautta. Tutkimuksessa uusien sovellusalueiden etsiminen on sisältänyt uusien sovelluskonseptien suunnittelun toteutuksen. Joissakin toteutetuissa tapauksissa vuorovaikutus jäljittelee ihmisten välistä vuorovaikutusta, kun toisaalla vuorovaikutus ei pohjaudu fyysisiin esikuviiin.

Ensimmäisessä sovelluksessa - DIVA virtuaaliorkesterissa - kapellimestari voi johtaa virtuaalisia muusikoita. Sekä ääni että muusikoiden animaatio syntetisoidaan tosiajassa tietokoneella. Järjestelmän tärkeimpiä osatehtäviä on tulkita kapellimestarin liikkeitä ja kontrolloida synteisiä sen pohjalta. Väitöskirjassa esitellään tarkoitusta varten kehitetty järjestelmä. Järjestelmä pystyy myös tunnistamaan kapellimestarin liikkeistä tunneinformaatiota. Alkuperäinen järjestelmä käytti kaupallista liikkeenseuranta-järjestelmää, joka on sekä kallis että helposti hajoava. Näiden puutteiden korjaamiseksi kehitettiin kiihtyvyyssantureihin perustuva liikkeenseurantamenetelmä.

Kapellimestariseurannassa ja muissa elekäyttöliittymissä ohjelman pitää muokata mittalaitteesta saatava raakadata käyttökelpoisempaan muotoon. Tätä varten kehitettiin uusi menetelmä: Flexible User Input Design (FLUID). FLUID-komponentin tarkoitus on helpottaa harvinaisten syötelaitteiden hallintaa ja tukea signaalinkäsittelyä ja -analyysiä.

FLUID-järjestelmä kehitettiin keinotodellisuuspohjaisia AnimaLand ja Helma -piirto-ohjelmia varten. Näitä ohjelmia silmällä pitäen kehitettiin myös uusia graafisia menetelmiä. Perinteisiä partikkelisysteemejä parannettiin lisäämällä liikkuvia voimakenttiä jotka vaikuttavat toisiinsa. Toisiinsa vaikuttavat voimakentät tekevät animaatiosta eloisampaa ja uskottavampaa.

3D grafiikan piirron hajautuksesta tuli ongelma, kun haluttiin käyttää useampaa tavallista tietokonetta 3D grafiikan piirtämiseen. Tätä varten kehitettiin menetelmä joka pohjautuu nopean paikallisverkon broadcast-teknologiaan. Menetelmä vähentää lähetettävän data määrää ja siten parantaa järjestelmän suorituskykyä.

Ääni on oleellinen osa monissa multimodaalisissa käyttöliittymissä. Tarkoitusta varten kehitettiin yleiskäyttöinen Mustajuuri-ohjelmisto. Mustajuuri on joustava ja tehokas äänenkäsittelyjärjestelmä, joka tukee erityisesti äänenkäsittelyä keinotodellisuusympäristöissä.

Avainsanat elekäyttöliittymät, kapellimestariseuranta, keino-
todellisuus, digitaalinen taide, grafiikkaklusterit, par-
tikkelisysteemit, 3-ulotteinen ääni, digitaalinen sig-
naalinkäsittely

PREFACE

Anna Nerkagi lives in the village Laboravaja in the northern Siberia. In the winter the temperature goes down to -50 degrees centigrade. In the summer the snow barely melts. At the end of the summer only the top 30 centimeters of the earth is molten. This is the home of the Nenets people. They are nomads, taking care of their reindeers and dogs in these harsh conditions.

Anna is the spiritual leader of Laboravaja and she has founded a voluntary hut-school in the village. In this school the teaching is based on three central themes: Joy, How to do good and Nenets tradition. Anna tells us why these things are important: Joy makes life worth living. Doing good things makes joy for oneself and others. Children who know the Nenets tradition know who they are, who their parents were and where they live.

Today only a fraction of the Nenets people live in villages like Laboravaja. There is a great risk that once Anna dies the village will disappear, like the whole Nenets culture has been disappearing since Stalin's reign. Most Nenets live in arctic cities where the most common cause of death are suicide and alcoholism. The Nenets way of life, tradition, art and language will disappear during the next 50–100 years. Anna knows all of this, but she still has the wisdom and the strength to teach Joy, How to do good and the Nenets traditions.

What can we learn from her? This lesson is simple because it can be understood with a few words. It is also extreme because the issues are present in all human activity and because her example sets a high standard on what a human being it meant to accomplish.

This thesis makes a modest attempt to learn something from Anna Nerkagi and to apply the same principles in a work that is carried out in the realm of modern science. Scientific rhetorics is possibly the weakest medium for discussing Joy, How to do good or Who we are. Hopefully these themes can be seen in some form on the following pages.

Otaniemi, Espoo, 8.9.2006, On the feast of nativity of Theotokos, the Mother of God

Tommi Ilmonen

ACKNOWLEDGMENTS

This thesis has been done in the Helsinki University of Technology (HUT) Telecommunications Software and Multimedia Laboratory (TML) between 1999 and 2006. The work is related to my earlier work in the same place from 1996 to 1999.

I am thankful to all people in the EVE-team in TML. In particular to Tapio Takala (advisor of this thesis and a leader of the digital media group), Lauri Savioja (co-leader of the media group in TML) and Wille Mäkelä (leader and initiator of the Helma project).

Other notable colleagues in TML have been: Timo Aila, Juhani Forsman, Matti Gröhn, Jarmo Hiipakka, Rami Hänninen, Liisa Hirvisalo, Perttu Hämäläinen, Janne Jalkanen, Ipo Lahtinen, Raine Kajastila, Aki Kanerva, Vesa Kantola, Petteri Kontio, Janne Kontkanen, Juha Laitinen, Jaakko Lehtinen, Tapio Lokki, Janne Nummela, Teemu Mäki-Patola, Hannu Napari, Markku Reunanen, and Seppo Äyräväinen.

Less direct colleagues have been Anu Kankainen, Tomi Kankainen, Karri Palovuori and Jari Takatalo.

This work has been funded by the Helsinki Graduate School on Computer Science and Engineering of the Academy of Finland (HeCSE). During the Helma project I have received funding from Tekes. The final stages of the thesis work have been supported by Helsinki University of Technology and the the Finnish Cultural Foundation.

Finally I am deeply indebted to Kata Koskivaara, Aile-Matleena Ilmonen and Pekko-Hermann Ilmonen for love and support.

TABLE OF CONTENTS

Abstract	1
Tiivistelmä	3
Preface	5
Acknowledgments	7
Table of Contents	9
List of Publications	11
List of Abbreviations	13
1 Introduction	15
1.1 Scope of this Thesis	15
1.2 Research Questions	15
1.3 Overall Research Structure	16
1.4 Organization of the Thesis	16
2 Motivation	17
2.1 Visions	17
The DIVA Band	17
Storm	18
AnimaLand	18
Helma	19
Upponurkka	19
Kylä	21
2.2 Philosophy	22
Multimodal Interaction	22
Physical Interaction	23
The Idea of Man	23
Of Man and Technology	24
2.3 Technology	25
Input Device Management	25
Gesture Analysis	25
User Interfaces for Physical Interaction	25
Graphics Tools	26
Audio Tools	26
3 Related Research	27
3.1 Multimodal User Input	27
3.2 Gestures in the User Interface	27
Conductor Following	28
3.3 Emotions from Motion	29
3.4 Work in Virtual Reality	29

Immersive Graphics Tools	31
3.5 Graphics Systems	31
Particle and Fluid Effects	31
Rendering Clusters	32
3.6 Audio Processing	33
4 Flexible User Input Design	35
4.1 System Requirements	35
4.2 Selected Techniques	35
4.3 Discussion	36
5 Physical and Multimodal User Interfaces	39
5.1 Conductor Follower	39
Motion Tracking	40
Tempo Analysis	42
Nuances	43
Emotion analysis	44
Player Modeling	46
DIVA System in Use	46
5.2 Helma	47
Virtual Pockets	48
Upponurkka	49
5.3 Kylä	51
5.4 Discussion	52
6 Techniques for Audiovisual Effects	55
6.1 Mustajuuri Audio Engine	55
Discussion	56
6.2 Graphics Techniques	57
Second Order Particle Systems	57
Broadcast GL	58
Discussion	59
7 Future Directions	61
8 Summary	63
9 Main Results of the Thesis and Contribution of the Author	65
Bibliography	67
Errata	79

LIST OF PUBLICATIONS

This thesis summarizes the following articles and publications, referred to as [P1]-[P8]:

- [P1] Ilmonen, Tommi and Kontkanen, Janne. Software Architecture for Multimodal User Input – FLUID. In Universal Access. Theoretical Perspectives, Practice, and Experience: 7th ERCIM International Workshop on User Interfaces for All, Lecture Notes in Computer Science 2615, pages 319–338, Springer Berlin / Heidelberg, 2003.
- [P2] Ilmonen, Tommi and Takala, Tapio. Conductor Following With Artificial Neural Networks. In Proceedings of the International Computer Music Conference, pages 367–370, Beijing, China, 1999.
- [P3] Ilmonen, Tommi and Jalkanen, Janne. Accelerometer-Based Motion Tracking for Orchestra Conductor Following. In Proceedings of the 6th Eurographics Workshop on Virtual Environments, Amsterdam, Netherlands 2000.
- [P4] Ilmonen, Tommi and Takala, Tapio. Detecting Emotional Content from the Motion of an Orchestra Conductor. In Gesture in Human-Computer Interaction and Simulation, 6th International Gesture Workshop, Lecture Notes in Artificial Intelligence 3881, pages 292–295, Springer Berlin / Heidelberg, 2006
- [P5] Ilmonen, Tommi and Reunanen, Markku. Virtual Pockets in Virtual Reality. In Virtual Environments 2005 Eurographics/ACM SIGGRAPH Symposium Proceedings, pages 163–170, 2005.
- [P6] Ilmonen, Tommi. Mustajuuri - An Application and Toolkit for Interactive Audio processing. In Proceedings of the 7th International Conference on Auditory Displays, pages 284–285, Helsinki, Finland, 2001.
- [P7] Ilmonen, Tommi and Kontkanen, Janne. The Second Order Particle System. The Journal of WSCG, 11(2):240–247, 2003.
- [P8] Ilmonen, Tommi, Reunanen, Markku, and Kontio, Petteri. Broadcast GL: An Alternative Method for Distributing OpenGL API Calls to Multiple Rendering Slaves. The Journal of WSCG, 13(2):65–72, 2005.

LIST OF ABBREVIATIONS

<i>API</i>	Application Programming Interface
<i>ANN</i>	Artificial Neural Network
<i>AR</i>	Augmented Reality
<i>BGL</i>	Broadcast GL
<i>CAD</i>	Computer Aided Design
<i>CAM</i>	Computer Aided Manufacturing
<i>CAVE</i>	CAVE Automatic Virtual Environment
<i>DIVA</i>	Digital Interactive Virtual Acoustics
<i>DSP</i>	Digital Signal Processing
<i>EVE</i>	Experimental Virtual Environment
<i>FLUID</i>	FLexible User Input Design
<i>GUI</i>	Graphical User Interface
<i>HCI</i>	Human-Computer Interaction
<i>HMD</i>	Head-Mounted Display
<i>HRTF</i>	Head-Related Transfer Function
<i>HUT</i>	Helsinki University of Technology
<i>IP</i>	Internet Protocol
<i>MIDI</i>	Musical Instrument Digital Interface
<i>MLP</i>	Multi-Layer Perceptron
<i>OpenGL</i>	A graphics API for real-time computer graphics
<i>PC</i>	Personal Computer
<i>Pd</i>	Pure Data, an application for data-flow processing
<i>SGI</i>	Silicon Graphics, Incorporated
<i>SOM</i>	Self-Organizing Map
<i>TCP</i>	Transport Control Protocol
<i>TML</i>	Telecommunications Software and Multimedia Laboratory
<i>UDP</i>	User Datagram Protocol
<i>VBAP</i>	Vector Base Amplitude Panning
<i>VE</i>	Virtual Environment
<i>VRPN</i>	Virtual Reality Peripheral Network
<i>VR</i>	Virtual Reality
<i>X11</i>	The graphical windowing system used by most UNIX platforms
<i>3D</i>	Three-Dimensional

1 INTRODUCTION

This thesis concerns design, realization and testing of several multimodal interaction systems. A general goal of this research has been to find new applications for multimodal interaction.

1.1 Scope of this Thesis

This thesis is defined by interaction concepts (or visions) that have driven the research. These concepts are related to each other by common technology and background philosophy.

All of the concepts share some user interface components. The components fall into several categories. As far as user input is concerned many concepts share similar needs for input data management. This has led to research on how to handle multiple input devices. Once the data is collected it needs to be analyzed. Here an orchestra conductor follower serves as a case.

On the output side many concepts use audio and graphics. The audio feedback has resulted in the development of a flexible audio processing engine. The graphics synthesis has involved particle systems and the creation of a 3D drawing application. For the drawing application we also needed new user interface concepts, which led to the inclusion of the virtual pocket user interface metaphor.

1.2 Research Questions

The primary research question has been to find out how multimodal applications could be used to create human-computer interaction systems that could not be realized with desk-top or mobile computing. This question is further split into two sub-questions:

1. What kind of technical tools are needed for the development of multimodal applications? This question asks what kind of tools there should be. To answer this question several software tools were constructed and used to build multimodal applications. The innovations and experience gained from these experiments provide the answer to the research question.

2. What kind of multimodal applications are useful? Since the design space for multimodal applications is extremely large one needs guidelines in concept and application design. Given the wide range of applications the goal needs to be realistic. In this case we looked into specific concepts and experience we gained is useful in areas close to implemented systems.

The two questions are directly related, which is the reason why they are taken up in the same thesis. This research is not expected to give a complete answer the questions, but shed some light on the problems and solutions that can be expected.

1.3 Overall Research Structure

This thesis represents open-ended, constructive research. The overall idea has been to explore the possibilities of multimodal interaction in different forms using a series of concepts that illuminate various parts the target field.

The development of multimodal applications constantly demands considerations about the application concept, user interface and available technology. This has called for multi-disciplinary research where the whole web of components has to be taken into account at the same time. This requires open mind and flexible targets since one is constantly seeking a compromise between the different requirements and limitations.

The human-centric concepts have often led to technological research that is necessary to realize the concepts. Construction of the new technologies then became the scientific contribution of this thesis.

1.4 Organization of the Thesis

This thesis is organized as follows. Chapter 2 describes the motivation and philosophical background for this work. In Chapter 3 related research is presented. Chapters 4–6 cover the technologies that were developed during this work. Chapter 7 extrapolates some future directions and Chapter 8 gives a final overview of the work. The scientific contribution of each research paper is summarized in Chapter 9.

2 MOTIVATION

2.1 Visions

There were several visions that drove this research and defined the specific research topics. During this research many of these concepts have been realized while some have been left unimplemented.

The DIVA Band

The Digital Interactive Virtual Acoustics (DIVA) -band was an effort to create a virtual orchestra performance. A real conductor would conduct a band of virtual musicians with a baton (figures 2.1 and 2.2). In the system multiple technologies were merged — inverse kinematics for calculating the motion of the musicians, physical modeling for the musical instruments and room acoustics and conductor following to translate the intentions of the conductor to MIDI commands.



Figure 2.1: A conductor with a tracked baton and glove.

The DIVA system was showcased in the SIGGRAPH'97 conference in the Electric Garden [57]. This was a large installation with magnetic motion tracking, large video screen and head-related transfer functions (HRTF) filtering of the final audio signal.



Figure 2.2: The DIVA virtual band.

After the SIGGRAPH debut DIVA band became the standard laboratory demo. We also made a low-cost version for the Heureka science center. For this setup we developed a motion tracking system that was based on accelerometers.

Tapio Takala is the author of the virtual orchestra concept. The author's role was to implement the conductor following software.

Storm

Storm was an installation concept by the author. In the Storm three users are located in front of video screens — one screen for each user. Their goal was to use gestures and voice to control the weather in a virtual landscape. For example one of the users could be wind while the other was rain. With the installation the users would create different atmospheric, aesthetic effects together. This effects and (possibly supporting a story-line) would be the content of the installation.

The storm concept was never fully realized. As a first step towards it we built a particle animation application — AnimaLand — that was to be a simple test-bed for many of techniques in Storm [65].

AnimaLand

To prepare for the Storm installation we began work on some techniques that were necessary for the Storm. As a first step we planned to create a particle animation application — AnimaLand. AnimaLand was to be a virtual reality (VR) application where an artist could create animations in real-time with gestures (figure 2.3) [65].

AnimaLand was aimed to demonstrate a series of software components that would make their way to the Storm. These were multimodal user interface with the Fluid [P1] library as input the interface, audio output with Mustajuuri sound processing engine [P6] and and graphics controlled by the VRJuggler VR framework [71].



Figure 2.3: The author is making a particle effect with the AnimaLand software.

The author is responsible for both the concept and implementation of AnimaLand. While AnimaLand was to be test-bench for Storm it was also designed to be an independent application.

Helma

The aim of the Helma project was to create a 3D drawing and animation environment in VR [96]. The special interest was to see how fine-motoric interaction affects the creation of digital 3D art (figure 2.4). The basic assumption behind the project was that enabling physical interaction would create digital art with visible human touch, like in traditional paintings. Artists would also be able to use their highly trained motoric skills to create digital 3D art.

Wille Mäkelä is the author of the Helma concept. The details of the user interface and interaction design were done by the Helma team: The author was responsible for almost all of the software design and implementation, Markku Reunanen took care of the hardware side, Wille Mäkelä for overall goal settings and user feedback and Tapio Takala for further ideas and guidance.

The software written for AnimaLand was taken as the starting point for the Helma software. This enabled us to use a working base system the starting point.

Upponurkka

The works created with Helma were presented with an immersive, stereoscopic large-scale installation called Upponurkka (immersion corner). The



Figure 2.4: An artist is drawing a tube with the Helma system.



Figure 2.5: The Upponurkka installation that was used to display the 3D art made with the Helma system.

system consisted of two large screens with polarization-based stereoscopic graphics. Number of people can view the graphics at the same time, but only one person has a pair of glasses that are tracked based on colored markers.

Upponurkka was part of the Helma project and the division of work is identical with the overall project work, outlined in the previous section.

Kylä



Figure 2.6: A visitor in the Kylä installation.

Kylä (village) is an interactive art installation by the author. This installation occupies one room. The visitors are given a small bee's-wax candle that is the only source of light in the room. There is only one candle in the room at a time (figure 2.6). On the walls there are photographs taken during the 19th century in the northern Carelia (an area currently shared between Finland and Russia). Each image is accompanied by a specific rune-song that is played as the candle approaches the picture. The songs and their melodies come from the same region as the photographs, but they are sung by modern singers. An exception to this rule is a photo of a graveyard. It is accompanied by the last song of the Orthodox funerals.

The aim of the Kylä installation is to capture some of this atmosphere and present it to modern people (kylä = village in Finnish). It serves as a document, presenting fragments of people's lives, beliefs and hopes. Components that make up the experience are a candle, a darkened room, photographs from the 19th century, folk songs from the same era and fire-based human-computer interaction. In this installation the computer is both physically and mentally hidden.

The installation was built early in 2005 when it was demonstrated for two weeks in a small media art festival in Helsinki University of Technology.

For the summer 2006 Kylä was in the Parppeinvaara open-air museum in Ilomantsi, eastern Finland.

The author created the installation idea which was further developed and implemented with Marja Hyvärilä. The rune-songs were sang by the author, Kati Heinonen, Taito Hoffren, Vilja Mäkinen, Tuomas Rounakari, Juulia Salonen and Veera Voima. The first version of the funeral song was sang by the author, Petri Hakonen, Stina Hakonen, Kati Heinonen and Jaakko Olkinuora. A more ethnographically accurate version of the song was arranged in Finnish by Tuomas Rounakari, based on Old-Believer recording and sang by the author, Charlotta Buchert, Anne-Mari Rautamäki, Niina-Mari Rautamäki and Tuomas Rounakari.

2.2 Philosophy

The visions above share a common philosophical background. The philosophical considerations fall into two major categories — multimodal interaction and physical interaction. All the visions contain both ingredients in some form.

Multimodal Interaction

A central theme in this thesis is multimodal interaction. The term “multimodal” has multiple meanings depending on the context. It can mean interaction with multiple devices (mice, keyboards, cameras) [139], the use of multiple human skills (writing, drawing, gesturing, talking) [4] or the use of different sensory channels [101]. The first definition is centered on technology while the two others are human-centric. In this thesis the term is used simultaneously in all meanings — a multimodal HCI system usually needs to use a number of different devices. At the same time multimodal applications require the use of multiple skills and sensory capabilities.

One weakness of the term multimodal is its volatility. In principle even a normal GUI application is multimodal — the communication between the human and application uses graphics, text, sounds, the keyboard and the mouse. A “multimodal” application uses more modalities than a normal GUI application.

Current human-computer interfaces use only a small subset of possible modalities, usually just text and images. Alternative communication channels cannot be used (speech, gestures, body-language). It is seen that the use of these alternative channels can improve the user interface [116]. Multimodal interaction offers several potential benefits over unimodal interaction.

First of all new modalities may offer new application areas for technology. There are situations where the traditional input methods simply do not work, for example few people draw comfortably with a mouse or even a drawing pad. There are also people who cannot cope with present interfaces (e.g. children and illiterate people). The question is about how well a given interaction style fits people. It is unreasonable to expect that the same user interface would be ideal for all users.

Secondly performing the same task in different way may have an effect on how the task is performed. Even though people seldom use computers

just for the sake of interacting with them the method of interaction needs to be considered carefully. If a multimodal interface is more pleasant it can contribute directly to better user experience. In some cases this experience may be one of the most important reasons to use a computer in the first place, in games for example.

Thirdly the optimal modality for interaction may depend on the context. For example the user interfaces in desktop computers, mobile devices and embedded systems are so different that one may be forced to use different modalities due to the different nature of the systems.

Physical Interaction

Physical user interfaces emphasize the role of the user's physical activity. Traditionally the user's body is seen mostly as a problem — e.g. countless studies and text-books analyze the limits of human vision, hearing and accuracy [107, 151, 117, 159]. In the realm of physical interaction the user's body is seen as an important part of the interaction [68, 38]. The aim of such thinking is to widen the scope of human-computer communication from mental models to psycho-physical aspects. In a more concrete form: Instead of activating just eyes, fingers and wrists (keyboards and mice) a human-computer interaction system should utilize the user's physical skills as much as possible.

An important topic is also the direct impact of physical exercise. Strong physical activity causes adrenaline flow which makes one feel better. This kind of biological aspects are one part of human-computer interaction. In classical HCI this part has received little consideration.

Physical interaction has also wider sociological links. In the developed countries most people do little physical work. Physical activity has become something special. Hobbies (e.g. sports and trekking) make a clear distinction between normal life and exercise.

Here it is useful to observe modern schools and office work. The modern man spends enormous amounts of time sitting down, looking at rectangular papers or computer screens. Pauses for physical exercise are allowed only because without them people's mental productivity suffers. Physical activity is tolerated, only because it cannot be completely removed.

The Idea of Man

Often multimodal and physical interaction are seen as methods to break from traditional HCI. It is seen that typical HCI is concerned almost entirely with efficiency. For example in the 1990's Picard criticized modern HCI and proposed that affective computing would make the modern technological world better for humans [119, 118]. At the core of Picard's criticism is the belief that traditional HCI (and science in general) ignore a vast area of human experience. Here Picard's thinking parallels Lakoff's [84]. As an empirical linguist Lakoff criticizes older objectivistic linguistic theories (e.g. Chomsky's generative grammar -theory) for assuming that language is a system of meaningless symbols. Backed by considerable empirical evidence Lakoff argues that natural languages are based directly on physical human experience. As a side effect the philosophical idea that language, logic and thought are universal should be abandoned.

Both Picard and Lakoff see science as a system where the thought/body duality is a dominant paradigm. This assumed duality is what led Descartes to declare *cogito ergo sum* (I think, therefore I am). The aim of Descartes was to establish the existence of pure reason. This pure reason would be able to create pure, universal knowledge that would be independent of our physical existence. Both Lakoff and Picard see this spirit in their respective fields of science as a problem. In similar way sociologist Norbert Elias notes that pure, incorporeal knowledge is only a philosophers' dream [42]. Also the phenomenological tradition of philosophy (Edmund Husserl, Martin Heidegger and others) have concluded that thought cannot be separated from the body.

It seems that the source of conflict between Descartes and the above scientists and philosophers is in their different idea of man. The western, rationalistic approach has been to split problems into parts. When a human is perceived as parts the mind and body that are handled separately. Furthermore the mind is divided into rational and irrational parts. In multimodal interaction design these assumptions are typically rejected. It seems that most people who make physical/multimodal systems approach the user holistically, assuming that the physical and mental activities are directly coupled.

Of Man and Technology

Since we are dealing with both the idea of man and technology it makes sense to shortly discuss the relationship between humanity and technology. Human is basically a cultural animal — a person's behavior is shaped by the culture he/she lives in. Word culture is used here in its ethnographical meaning: Culture is the part of human inheritance that is not biological. Technology is one aspect of this human heritage. The modern social order is based on massive technological infrastructure. For example without effective transportation of goods and people large societies and high civilization could not be maintained.

Technology has freed humanity more and more from the limitations of environment, body and mind. Given the ease of survival and comfort of our lives it is a surprise that modern people are not as fit as their predecessors were. Reading and writing skills have made good memory obsolete — people in illiterate societies tend to have better memory and different brain structure [97]. The elaborate strategies of the antique for memorizing things have also largely disappeared.

There are powerful commercial and political interests in technology. Large corporations and nations try to modify life-style and legislation worldwide to a direction that is beneficial to them. Technology is one venue where such push can be made.

Given these facts it seems that technology can affect people's thoughts and behavior to a great extent. While technology is only one factor in society, it is the domain of this work. Legislation, culture and fashion etc. are also important (possibly more so), but they are outside the scope of this thesis.

2.3 Technology

Above we have outlined the visions we had and their background. In this section we cover the technologies that were needed to realize the visions. These new technologies form the core of this thesis.

Input Device Management

Input device management is one of the basic problems in multi-modal interaction. In the realm of graphical user interfaces one can use a common widgets/callbacks -architecture (e.g. Qt [70] and Motif [17]). There are many toolkits that implement this common approach. This helps application development since individual projects do not need to re-create the same tools. In multimodal interaction such standards do not exist yet. Neither in the conceptual (common architectures and metaphors) or practical level (common toolkits).

As a solution to these problems we developed the FLEXible User Input Design (FLUID). This architecture is suitable for any application where novel input devices are in use. The system is scalable from embedded systems to ordinary computers. The design takes into account the needs of higher-level application development — support for input data processing (gesture detectors etc.) and ease of programming. While the system is generic in nature we have developed and used it primarily in virtual reality (VR) applications.

Gesture Analysis

If we want to create a gesture-aware application these gestures need to be detected first. Orchestra conducting is one case where gestures are used to convey meaning between humans. This is a challenging task since there is no standard conducting language, but each conductor has his/her style. Humans understand these individual styles naturally, while a computer has to be instructed to understand every motion.

To mimic this interaction we built a conductor following system that extracted conducting gestures from the conductor's motion and synthesized music as required. The software detected tempo, dynamics and emotional content from the conductor's motion. In the music synthesis we did not utilize the emotional information, but all other information was used.

User Interfaces for Physical Interaction

One of the problems in VR/AR interaction are the interaction techniques (or metaphors). Instead of using common techniques each application uses a different approach for application control. While such specialization might lead to very effective user interfaces it seems that mostly new techniques only provide a slightly new way to perform some user interface task (e.g. tool selection and configuration).

This was also a problem in the Helma project. One of the problematic parts was the tool selection. Our first attempts at tool selection and configuration were clumsy and they did not support artists' work-flow. We adopted the virtual pocket metaphor for tool selection needs. In user tests this proved

to be an effective method.

Graphics Tools

The Storm, AnimaLand and Helma concepts required novel graphics tools. For the Storm and AnimaLand we needed particle systems. Traditional particle systems were considered too static for the purpose. To create more dynamic effects we developed the second order particle system.

In projection-based VR we have to display the same virtual world on multiple display surfaces. The traditional approach to meet this need has been to purchase a computer that is capable of running several data projectors at once. Unfortunately such computers are expensive. Due to cost constraints we wanted to use off-the-shelf components instead of costly special hardware. In practice this implies that we use a cluster of ordinary PCs to render the multiple screens. For the clustering we needed a graphics distribution method.

Previously there have been two ways to distribute the graphics: application distribution and distributed rendering. The first offers usually better performance, but it takes more work. The latter is easier to make but places heavy stress data transfers between the application and rendering computers. As a method to reduce this data traffic we created the Broadcast GL toolkit (BGL). This is a system that uses networking resources as efficiently as possible.

Audio Tools

A multimodal application often needs an audio engine. To work well in a VR installation the engine must be able to produce sound at low latency, multiple sound synthesis methods and 3D panning of sound sources.

We created a general-purpose audio toolkit to meet these needs. This toolkit — Mustajuuri — offers a flexible environment for sound signal processing.

3 RELATED RESEARCH

Given the multidisciplinary nature of this research this chapter covers relevant related work in sections that match the topics of the thesis publications.

3.1 Multimodal User Input

Multimodal user input is a field that can be approached from many directions. Nigay et al. have made such high-level analysis [89]. Their emphasis is on the classification of different interaction modes and modalities. Schoemaker et al. have also published work on the levels of observation [136]. Their work classifies four levels of observation – physical/physiological, information theoretical, cognitive and intentional.

In addition to high-level considerations there are several toolkits to handle user input. For example Salber has published a “The Context Toolkit” for sensing the presence of the user [131]. This toolkit has been used by e.g. Mankoff who combined it with speech recognition [98]. The system was used to collect and process ambiguous user input data.

Many applications also process the input data to extract interesting non-obvious information from it. For example Landay has created the SILK-library for 2D sketch making [86]. Cohen has done research on an ambitious distributed input data collection and processing system, QuickSet [27].

In the VR environments one is typically confronted with non-standard input devices (motion trackers, data gloves etc.). Most VR frameworks have some method of handling these devices, for example VR Juggler [71, 7] and CAVELib [25]. Goldiez et al. have also developed a software infrastructure for multi-modal interaction [51]. These are monolithic VR systems that do not fit the needs of people who wish to make interaction systems outside VR environments. There are also VR device toolkits that contain only the device management, without other VR framework i.e. OpenTracker [126] and VRPN (virtual reality peripheral network) [148]. Bimber has worked on a multi-layered architecture for sketch-based interaction within virtual environments [9]. Latoschik has developed a system that performs gesture detection and combines the results with speech input [88]. These user input systems are either application- (e.g. sketching) or domain-specific (e.g. VR). They also tend to have be tied directly to the application logic, which is problematic if the envisaged logic does not fit the real usage.

3.2 Gestures in the User Interface

Gestural user interfaces offer an alternative to the traditional human-computer interaction methods. First systems were made in the 1970’s. Myron Krueger’s VIDEOPPLACE experiments are influential in this field [82]. In these experiments Krueger tested different interaction paradigms in the form of games or art.

There are many different motivations for gestural interfaces. Hämäläinen et al. have reported Kick-Ass Kung-Fu — an artificial

reality martial arts game [54]. In this case gestural communication was motivated by the desire to make a user interface that enabled one to practice martial arts against virtual opponents. Keates and Robinson have developed gestural user interfaces for the disabled [76]. In this case the traditional mouse/keyboard interfaces were rejected since they could not be used by the handicapped people.

There are also other cases where the the keyboard and mouse cannot be used. Waldherr et al. have used gestures for human-robot interaction [157]. Wu and Balakrishnan used hand- and finger gestures for tabletop interaction [165], while Grossman used similar interaction with volumetric 3D displays [52].

In the field of computer graphics gestural, multimodal interaction was pioneered by Bolt [10]. His visionary “Put-That-There” system combined speech and gestures to create a more natural user interaction modality. This work has been continued by e.g. Billingham [8]. Nielsen has studied how to develop intuitive gesture interfaces [113].

All in all gestural and multimodal interaction are topics with plenty of activity, but no dominant paradigm. Even the “Put-That-There” -approach remains a novelty 25 years after its creation.

Conductor Following

Orchestra conductor’s motion is a special-purpose gesture language and an interesting test case for natural gestural interaction. Conducting guides give detailed information on the language and its use (for example McElheran [105] and Rudolf [130]). In practice each conductor has his/her own style that may be far from the text-book style. One of the greatest challenges for a conductor follower is to be able to understand conductors regardless of their personal differences. This is basically a language understanding task that can be demanding for real musicians as well.

The goal of a conductor following system is to extract meaning from conductor’s motion and to react musically. Conductor following has been studied since the 1970’s. Pioneering work was done by Max Mathews, who developed the Conductor program [103] and the Radio Baton system [102, 12, 11]. Over the years numerous other systems have been built by Morita et al. [108, 109], Brecht and Garnett [18], Tobey et al. [149, 150], Usa and Mochida [152], Segen et al. [137], Murphy et al. [110] and Kolesnik and Wanderlay [81]. In addition to these conductor following systems Garnett has made a conductor training environment [50, 49].

While there are many published systems it is difficult to tell how much progress has been made over the years. Although modern approaches use more sophisticated algorithms this does not automatically lead to better performance. To verify the superiority of new systems one would need benchmarks that are unfortunately missing.

Marrin and Picard have developed special devices for conductor following, the digital baton [99] and the Conductors Jacket [100]. In Marrin’s work the emphasis has been on the development of the devices rather than on conductor following. These systems do not try to mimic orchestra’s reaction to a conductor but they have their own way of mapping device input to musical parameters. In practice this means that one can “conduct” a multitrack

sound file and alter the volume of the individual tracks without changing the tempo. Similar device-centric work has been carried out by Haflich and Burns [53].

The most important musical parameter that the conductor following systems control is the tempo. Some systems also follow dynamics (volume) and nuances (staccato, legato etc.). A real musician has more expressive control over the music, for example he/she can control the emotional content of the music. An advanced conductor follower should perform this task as well. Synthesis of emotional expression has been studied in length by Bresin and Friberg [47, 19, 20].

Conductor following serves as test-bed for gestural human-computer interaction. Conductors are experts in gestural communication, which usually guarantees that they can produce gestures that are easy for humans to understand, even with little practice.

3.3 Emotions from Motion

While emotional man-machine interaction is an active field, computational analysis of the emotions conveyed by human motion is an understudied topic. Greatest effort has been on detecting emotions from facial expression and/or speech [134, 37, 166, 26]. Physiological signals are also a data source for analysis, as in Healey's and Picard's work [56].

Most active research on gestural emotion analysis has been carried out in the University of Athens, where Drosopoulos, Karpouzis, Balomenos and others have studied the field [39, 73, 6]. They have studied gesture recognition with the aim of connecting specific gestures to emotions. The underlying assumption is that emotions are expressed by specific gestures, for example putting hands over head represents one emotion. In our research we make a different assumption: emotions are not expressed in what movements are made, but how they are made.

In conductor motion the explicit gestures are dictated by musical parameters. How these gestures are made implies the emotions. The question is not about what gestures are made, but how they are made. There is some research on this field as well. So far the research has concentrated on understanding how humans perceive emotions from gestures, music and speech. For example Vines et al. have studied how people perceive emotions from musical performances [155]. Topic has also been studied by Dahl and Friberg [32]. This kind of research methods could be used to benchmark our emotion-detection systems against human subjects.

3.4 Work in Virtual Reality

Many of the systems in the thesis work in virtual reality. There are many ways to define virtual reality, for example Burdea has given the following definition [23]:

Virtual Reality is a high-end user-computer interface that involves real-time simulation and interaction through multiple

sensorial channels. These sensorial modalities are visual, tactile, smell and taste.

As can be seen from the above definition multimodality — “multiple sensorial channels” — plays an important role in VR (although the quote above ignores sound). For this reason VR is also a practical test-bed for multimodal interaction — the central topic of this thesis.

VR is an area that has been gaining ground since the early 1990’s. Though the early hype has vanished the branch has kept growing steadily, according to Brooks and Burdea [22, 23]. VR covers a wide range of systems and techniques, from desk-top systems to expensive VR installations with multiple screens and projectors. As the large VR systems are quite unpopular today it needs to be noted that at least CAVE-like technology is not a growing field. The work in this thesis has been carried out in the Experimental Virtual Environment (EVE) at the Telecommunications Software and Multimedia Laboratory (TML) at Helsinki University of Technology (HUT) [69]. EVE is a CAVE-like [28] installation with four display surfaces. Overview of the system is in figure 3.1.

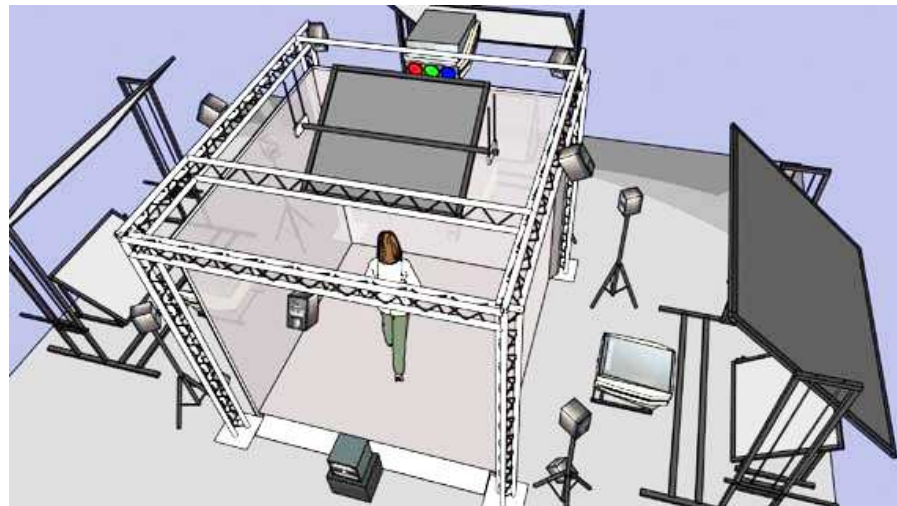


Figure 3.1: The Experimental Virtual Environment (EVE).

The systems in this thesis contribute to the understanding of human-computer interaction (HCI) in multimodal interaction, often in a VR environment. There has been little research on human factors in VR. Considering that user interface is a critical part of VR applications this is surprising. Most thorough research has been by Kaur whose PhD. work concentrated on usability in VR [74]. Her work concentrated on the high-level usability of complete VR systems. Kaur gives guidelines for developing usable VR applications. In similar spirit Knöpfle has described overall guidelines for improving usability in VR [80]. While he presents these as VR guidelines they are in fact generic guidelines for any HCI system. Gabbard et al. [48] and Bowman et al. [13] have studied different usability evaluation techniques in VR. Kalawsky has also written an overview of different evaluation techniques [72].

Most other research has concentrated on isolated tasks and methods without taking the big picture into account. One of the most common isolated research topics is navigation. Navigation is a fundamental element of interaction in VR. During the last decades several navigation systems have been proposed, but empirical research has been done by relatively few researchers e.g. Bowman et al. [16, 15], Usoh et al. [153] and Laakso [83]. Vinson has analyzed qualitatively how different landmarks can support navigation in VR [156].

Object selection and manipulation is another fundamental, isolated task that has been studied by many research groups, e.g. Bowman and Hodges [14], Márcio et al. [120], Mine et al. [106] and de Haan et al. [35].

When it comes to HCI, artistic work needs special attention. Simple performance metrics do not tell how well an application support artistic processes. For example flow is seen as an important factor in artistic processes and it should be included into usability research [29, 31, 30].

Immersive Graphics Tools

Art is an application area of immersive VR. One can use virtual reality as a pure art display medium (for example Davies' work [34, 33]). Another promising approach is to use VR as an art authoring environment. For example in 1988 AutoDesk founder John Walker concluded that virtual reality would be the next logical step in CAD/CAM user interfaces [158]^{page446}. In an immersive drawing/CAD/CAM application an artist would be able to draw directly into 3D space, thus bypassing the 2D mouse/keyboard interface.

The first immersive drawing system was reported by Wesche and Droske in 2000 [162]. Their system ran on the "responsive workbench". Since then others have made similar systems both in CAVE-like installations and on the responsive workbench. This work has been done by Keefe et al. [77], Li et al. [91] Schkolne et al. [135], Foskey et al. [45], Fiorentino et al. [44] and Mäkelä et al. [95, 96]. Mäkelä's articles refer to the Helma software (by the author), which is the only one to include animation. The Helma software is based on the AnimaLand (by the author), that was originally aimed to be a pure particle animation environment [64].

In general the systems have been developed based on the intuition of their makers. Sometimes the users' needs are different from the developers assumptions. For example Keefe reports that many of their drawing tools were rejected by artists [77]. To avoid such situations application development should be user-centric. As a support for user-centric design Deisinger et al. have studied the real needs of designers [36].

3.5 Graphics Systems

Particle and Fluid Effects

Particle systems are a method to simulate fuzzy objects and fluid phenomena. The technique has been in use already in 1970's, but the first scientific publications were made by Reeves in early/mid 1980's [124, 125]. More recently Sims has published detailed information on how to create particle effects [138]. Particle systems are currently part of main-stream computer

graphics, to the point where simple introductory papers have started to appear [87, 154, 104].

Flocks can also be seen as a special extension of particle systems, a flock member is particle with more complex behavior and appearance [128]. While particle systems can create powerful visual effects they are difficult to control. If the forces that create the animation are stationary the animations look monotonic. To counter this problem Wejchert and Haumann have developed a system where the forces could be moved easily [161]. This is the only publication that concentrates on making the particle systems easy to use. Since Reeve's original publication the basics of particle systems have remained the same. Modern particle systems may use greater number of particles with more elaborate behavioral rules, but there has been little progress in the fundamentals.

Particle systems can be used to create fluid effects. Another approach to fluid effects is the use of physical modeling. Jos Stam has worked on this topic for a long time [144, 141, 142, 143]. His work shows that fluid simulations can be carried out in real-time, for example for game applications. In addition to Stam Yngwe et al. [168], Fedkiw et al. [43] and Foster et al. [46], Lamorlette et al. [85] and Nguyen et al. [112] have studied fluid simulations. Many of these publications also target the creation and control of fire and explosion effects.

Rendering Clusters

There are two methods to display stereoscopic graphics. One is to use personal displays (HMDs etc.) and the other is to use a system with large display surfaces. Often such systems use more than one surface where the graphics are projected i.e. CAVE [28]. In this work we have used the latter approach. Originally both the Helma and AnimaLand system were developed on an SGI Onyx2 graphics computer. Unfortunately such computers are very expensive, making maintenance and upgrading costly. A normal PC would offer better price/performance ratio, but one PC cannot drive four stereo walls at one time. A solution is to use a cluster of PCs.

Such clusters are notorious for the programmer since the graphics need to be displayed on all nodes in tight synchrony. There are two main approaches to this problem. One is to distribute the application while the other is to distribute the graphics commands over a local-area network. The latter is more simple from the programmers perspective since the application does not need to take the clustering into account. For example the X11 windowing environment has support for distributing OpenGL API calls between computers [78]. The problem with this approach is that an application may make an enormous number of API calls, resulting in heavy data traffic between the rendering computers. Humphreys et al. have developed two API call distribution systems, WireGL [62] and Chromium [63]. To reduce network and rendering load both systems perform viewport culling before transmitting the graphics command stream. Chromium includes a stream-caching module to further reduce the network traffic [40]. Another method to minimize data traffic is to use network broadcast. This method has been used by Hewlett-Packard as described by Lefebre [90]. Unfortunately there is little information about the internals of the system or its real-world per-

formance.

Besides distributing the graphics API calls one can also distribute the application. This approach offers usually better performance at the cost of more complex programming work. There are also several projects to create clustering systems that would support the creation of distributed applications. For example the VR Juggler has been extended for cluster-based rendering [2, 115]. Other platforms have been developed by Winkelholz and Alexander [163], Schaeffer and Goudeseune [132], Yang et al. [167], Stoll et al. [145], Naef et al. [111] and Eldridge [41].

Overview of modern and upcoming VR system has been written by Streit et al. [146] and performance analysis of some modern systems has been done by Staadt et al. [140].

In cluster-based VR rendering is often only one part of the problem. Another is the frame synchronization between the computers. Especially active stereo system require that all display are updated at exactly the same rate (genlock). Since normal PC graphics cards lack hardware with proper genlock features this needs to be done in software. To solve this problem Al-lard et al. have developed “Softgenlock” that performs active stereo genlock on Linux systems [1, 3].

3.6 Audio Processing

In addition to graphics a typical multimodal application uses audio output. For this purpose most systems employ a generic audio engine that synthesizes the sound. Most commercial audio applications are too tightly tied to multitrack audio recording or processing and they cannot be used as real-time audio engines. Game audio engines in turn lack the capacity to run vector base amplitude panning (VBAP) on a large number of loudspeakers. Miller Puckette has developed the visual Max/Pd programming languages [121, 122]. The Max paradigm supports combination of different signal processing modules with a graphical user interface.

One of the earliest complete 3D audio system with was made in HUT/TML by the DIVA team [147, 59, 58]. The original system used HRTF panning with headphones to create a surround-sound illusion. Later the software has been extended to use the VBAP-method to distribute sound via multiple loudspeakers [123].

4 FLEXIBLE USER INPUT DESIGN

When one wants to create a multimodal application one of the first tasks is to capture multimodal user input data. Often this means that we must utilize novel input devices to be able to collect varying input data. Earlier in HUT/TML we dealt with each device using custom code in each application. To make work easier we wanted to use one input system in all applications. Lack of proper input device toolkits led to the development of FLEXible User Input Design (FLUID) toolkit (article P1).

4.1 System Requirements

When making the toolkit we defined a generic input device management architecture and implemented it. The basic design goals were ease of use, reliability, extensibility and modularity. By easy to use we mean that the system should be simple enough to be understood quickly. It also implies that the toolkit should be adaptable to different application structures. Reliability means that the toolkit should work correctly in all situations. Also the toolkit should not degrade incoming data in any way. Extensibility is needed since we cannot know in advance what kind of devices people want to use with their applications. Modularity helps to keep the system organized. It is also helps if someone needs to use only a part of the system.

4.2 Selected Techniques

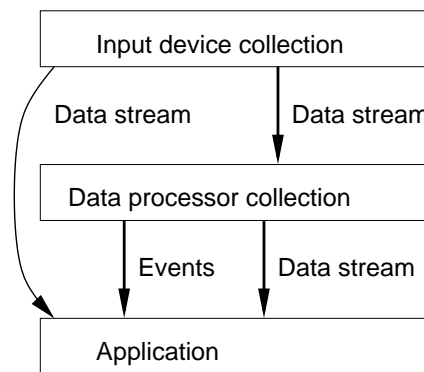


Figure 4.1: Overview of the FLUID architecture. The data streams from the input device collection are identical to the streams inside and from the data processor collection.

Figure 4.1 shows an overview of the FLUID architecture. FLUID contains two layers, the input device layer and the data processor layer. The input device layer manages all input devices. Each input device is managed by a specific object. Each device maintains a buffer of input data. The structure of an input device object is outlined in Figure 4.2. Each object

uses a separate thread to handle the actual device access. The incoming data samples are stored to an internal buffer. The application can call the input devices to update their data at any time. When this happens the data from the internal buffer is moved to a history buffer. This way the input devices can be updated even if the application stalls for some reason. If the buffers are long enough it is guaranteed that no input samples are lost. The latency of the system does not increase since the newest samples can be made available at any time with the update-call. From the application programmers perspective the system is single-threaded while it gets the benefits of being multi-threaded.

The raw input data is seldom useful. An application is typically interested in high-level features or events that need to be extracted from the input data stream. This work involves at minimum conversion of the input data from one format to another. Sometimes one also needs to use gesture recognition to detect interesting events. To support these needs FLUID has the data processing layer. Within this layer the application programmer can construct signal processing networks that together extract meaningful information from the input data. Typical input devices produce steady sample streams, while the application may be more interested in specific events (e.g. Figure 4.3). For this reason FLUID also supports event-based processing.

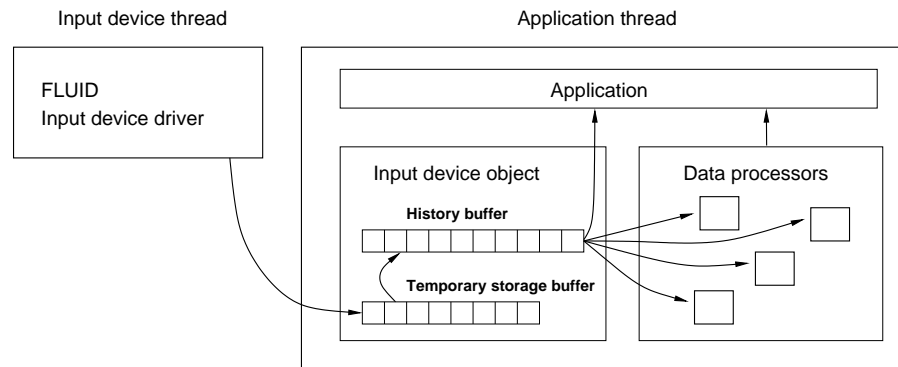


Figure 4.2: The input thread, buffers, and data transfer paths

The aim of the FLUID architecture is to provide a framework where programmers can create reusable components and share them between application. One could even write almost complete applications by combining proper data processing elements. In practice many applications have special requirements that need to be met. For this reason one can also bypass the data processing layer either partially or completely.

4.3 Discussion

FLUID was developed to be a toolkit that could be used to run the applications that were developed during the research. The need for such a toolkit was noticed after the conductor follower system was built. Since its creation FLUID has been used both in the installations of this thesis and in the work of others in HUT/TML.

Within the scope of multimodal user input processing FLUID offers a

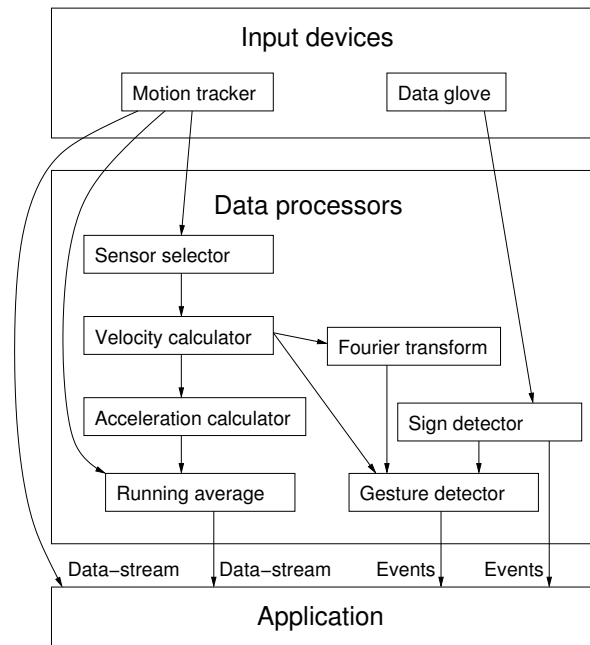


Figure 4.3: An example of how data can flow from input devices to the application.

unique combination of features. The selection of features was done to create the best compromise between the (often conflicting) requirements that such a system should fulfill.

FLUID uses a minimal design. It is more compact than many systems with equal features. Many systems use a multi-layer approach with different kinds of information processing taking place in different levels (e.g. Bimber's system [9]). The first FLUID design drafts included this kind of ideas, but they were later abandoned for several reasons:

- The layering is often artificial in the sense that the layers do not really represent the way the data is processed. For example, we could decide that one specific layer does semantic processing. While such labeling might make the system appear more structured it is likely that parts of the semantic processing are carried out in different levels anyhow.
- Such layering is typically application-specific and the toolkit should not dictate it in any way.
- Adding more layers to the system makes it more complex, which is clearly undesirable from both reliability and learnability perspectives.

FLUID makes little assumptions about what kind of data is passed inside the framework. This separates it from several other systems that are intended for very specific domains (e.g. OpenTracker [126] and VRPN [148]).

Each of the visions in Section 2.1 incorporates a novel user interface. This chapter covers the HCI -related topics of the visions.

The development of multimodal input management systems has proven to be a difficult task. Different applications require very different approaches, which makes the construction of general-purpose tools difficult. The FLUID architecture cannot solve this problem alone, but it can improve the awareness of the issues and provide working design patterns to projects that struggle with various input devices.

5.1 Conductor Follower

The conductor follower was part of a larger virtual orchestra project [147]. The DIVA virtual band combined multiple technologies from physical modeling of musical instruments to animation of musicians. One version of the system is shown in Figure 5.1. The complete DIVA system has been described in other publications [57, 92, 94].

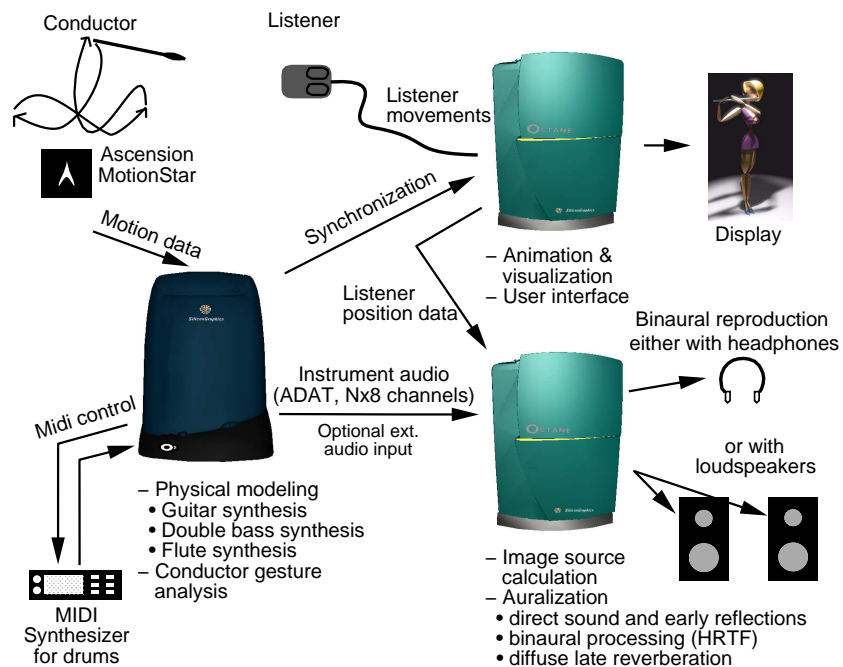


Figure 5.1: Overview of the virtual orchestra system in SIGGRAPH'97.

A conductor follower has several tasks to perform. First it needs to have access to the conductor's motion (motion tracking). After that it must analyze the conductor's gestures and extract musical information from the gestures. Then the follower needs to interpret the gestures in the current musical context. Finally it should synthesize musical reaction. These tasks depend on each other. For example knowledge of the current tempo and

dynamics can be used to aid analysis of the conducting gestures. Figure 5.2 shows an overview of the conductor follower and its components. Publication P2 gives an overview of the conductor follower system. The emotion analysis is described in publication P4.

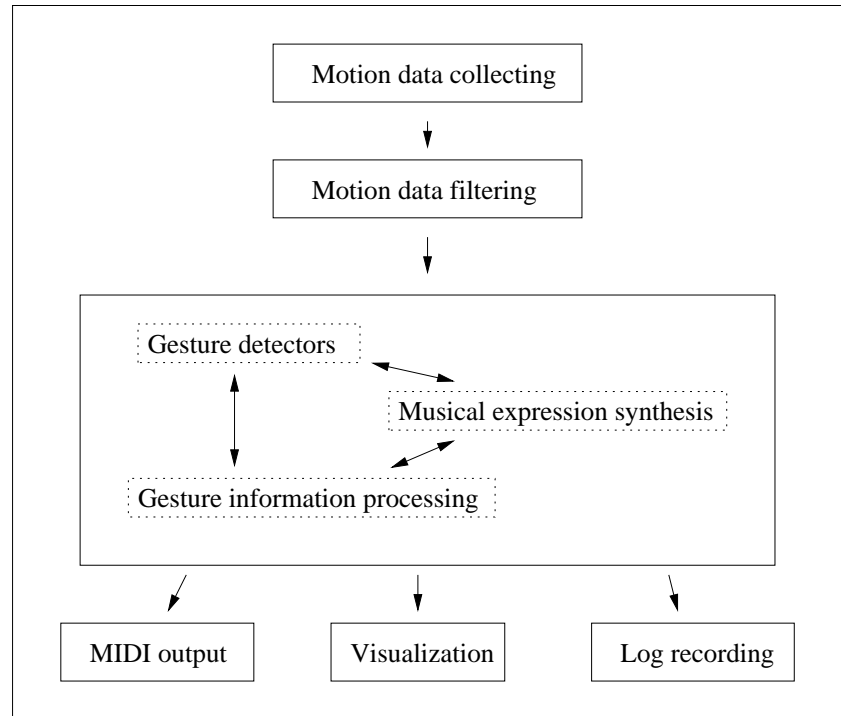


Figure 5.2: Logical structure of the conductor follower. The two two-way arrows indicate that the components that interpret the conducting motion influence each other.

Motion Tracking

At first we used the commercial Ascension MotionStar -system for motion tracking [5]. This system tracks the location and orientation of multiple small sensors. While the magnetic tracker is easy to use it was too expensive for a small-scale installation that we made to the Finnish science center Heureka. We were also concerned with the durability of the sensors. To solve this problem we began research on accelerometer-based motion tracker (publication P4). The aim was to create a cheaper and more durable motion capture system for the conductor follower. At the same time we could do research on the usefulness of acceleration tracking for conductor following. This approach is described in publication P3.

The principle of acceleration-based motion tracking is to measure the acceleration of some sensor and to integrate the acceleration twice to get location estimation. To track 3D motion a system with six sensors is needed. An inherent problem of this method is that any noise in the original measurement causes errors when integrating velocity from acceleration and location from velocity. This inevitably causes drift to the motion tracking system [21]. As a result the system needs to be calibrated frequently.

In conductor following this problem can be largely avoided. In this case the motion analysis works either with acceleration, velocity or relative location information. This is information that we can get reliably from accelerometer-based motion tracking system. We also know the limits of human movement and we can limit the motion estimations based on these limits.

In this work we used SCA 600 accelerometers by VTI Hamlin [55]. The motion tracking device consisted of three sensors mounted in one packet, as can be seen in Figure 5.3. This sensor packet was then attached inside the conducting baton. Only the vertical and the horizontal sensors were finally used for motion tracking, resulting in 2D motion tracking. For conductor following 2D motion tracking is typical sufficient because the movements form 2D patterns, when seen from the orchestras viewpoint.

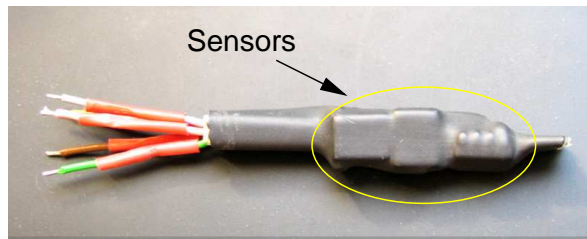


Figure 5.3: A prototype with three sensors mounted together.

While accelerometers always result in some tracking errors we can influence the kind of errors by choosing appropriate signal filtering methods. The signal flow from acceleration measurement to the motion estimation is in Figure 5.4.

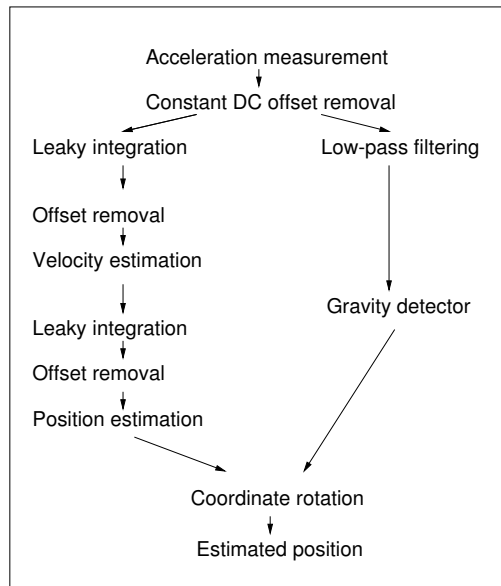


Figure 5.4: Signal flow in the acceleration-based motion tracking.

The signal processing has two major parts. The first part integrates the location information from the acceleration measurement values. It uses

leaky integration and dynamic offset removal (high-pass filter) to create zero-centered location estimation. This location estimation is in the coordinate system of the baton. To transform it into world coordinates we need to know the rotation of the sensors.

The baton can be picked up in any position, implying that each time the system is started there is new constant rotation. The constant rotation can be estimated by detecting the direction of gravity. In the acceleration measurements gravity appears as a constant offset into one direction. This direction can be calculated by low-pass filtering the acceleration measurements and using simple trigonometrics to calculate the rotation angle. Since gravity is always vertical we can use it to estimate baton rotation and transform the motion estimations from the baton coordinate system to world coordinates.

If the user rotates his/her hand while conducting this will cause a short-term error in the rotation estimations. Fortunately such rotations are limited by the conductor's physical capabilities and tendencies — we can assume that the baton rotation is more or less constant.

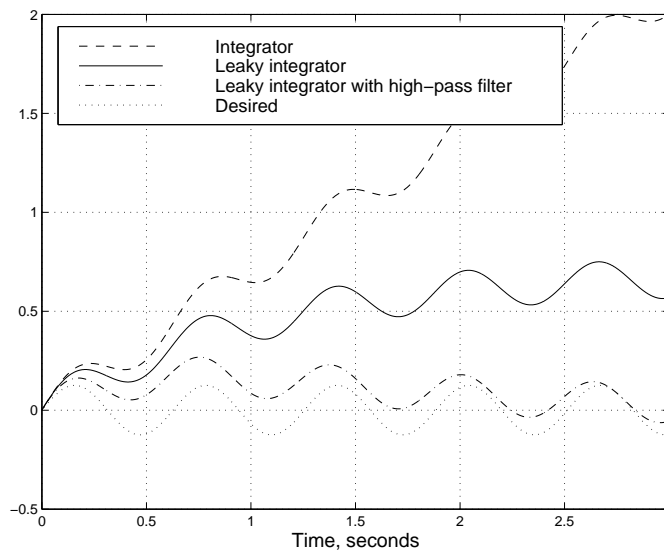


Figure 5.5: Using DSP to remove integration artifacts (numeric simulation). The resulting signal is slightly distorted.

Tempo Analysis

A conductor can communicate a wealth of information with his/her motion. Rhythm is usually the most explicit aspect of conducting and it is shown with a periodic, predictable motion. There are specific beat-patterns that are used with different time signatures. In practice conductors do not always follow the beat patterns exactly, as can be seen by comparing the ideal beat patterns in Figure 5.6 with real 4/4 beat pattern in Figure 5.7.

The basic tempo is determined by detecting down-beats from the motion. This is a common method, but it cannot react quickly to tempo changes, since we need to get at least one new beat to be able to determine new tempo. This lack of predictive capability led us to develop a beat phase estimator. The beat phase estimator is based on multi-layer percep-

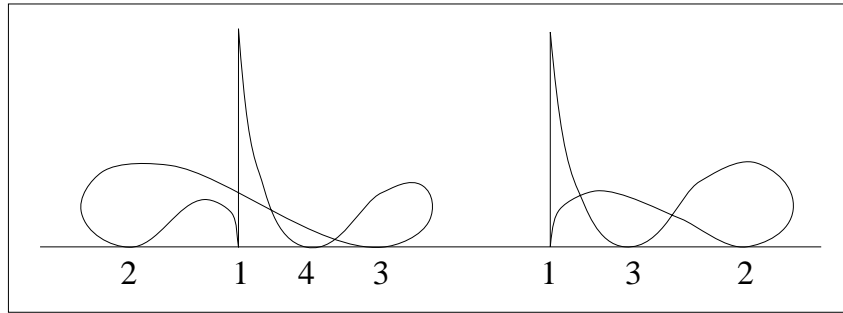


Figure 5.6: Conductor's view of beat patterns for time signatures 4/4 and 3/4.

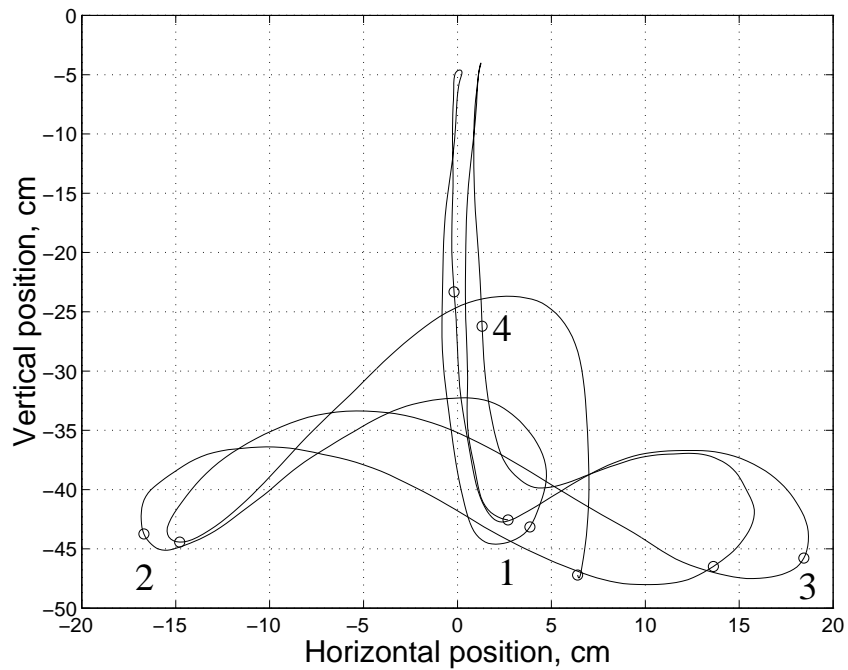


Figure 5.7: Measured baton tip movement for two bars of 4/4. Conducted very carefully by conductor Hannu Norjanen. The little circles indicate the point of curve where musical beats are supposed to be.

tron (MLP) artificial neural network (ANN) architecture [133]. As input to the ANN we feed current and past location and motion values. Based on these parameters the ANN estimates the beat phase.

These two approaches (base tempo and beat phase) yield two versions of the tempo information. The values are combined later when the system synthesizes the music.

Nuances

In addition to tempo a conductor can also control the nuances of the music. We have developed simple heuristics to control the volume and accentuation of the music.

The volume is controlled by two heuristics. The first is that large motion

implies large volume. The other rule is that rising the left hand with palm up is meant to increase the volume of the music, while lowering the left hand with palm down will reduce volume. The playback volume is affected by both heuristics simultaneously.

The system also handles staccato. A simple algorithm analyzes the acceleration of the baton tip and compares the maximum acceleration to the median acceleration. A great difference between these two values indicates that the motion is jerky and the follower should react by playing the music in staccato style.

Emotion analysis

Conductors can also express emotions as they conduct. The emotional cues are embedded into other conducting motions. Thus there is little point in trying to detect specific gestures for the emotions, rather we can try to find embedded features from motion. Publication P4 describes our method for detecting emotions from conductor motion.

As a starting point for the emotion analysis we used recorded conductor motion. The motion recording was done in a controlled environment where a conductor was asked to conduct passages of music with given emotional content. This setup demonstrates one fundamental problem of emotion research. The conductors are only acting some feeling, not necessarily feeling it. We cannot really know if the person is experiencing the feeling that he is trying to convey. As a solution to this it is sometimes proposed that one should use some biometric test to evaluate the true emotional state of the test subject. This would remove the need for the introspective aspects of the test. The problem with such approach is that the biometric test cannot be made without human evaluation, causing the same problems that it tries to solve. Additionally one would need to prove that the biometric test is correct in the context of conductor following. Due to these problems we declined to use biometric tests.

Given the importance of emotions in communication we wanted to test if one could extract the emotional content from the motion with computational methods. Our emotion detection system works in two phases, first we calculate feature vectors from the conductors motions and then apply artificial neural networks to analyze the feature vectors. The feature vectors should pick universal properties from the conductors' motion. To accomplish this the preprocessing phase uses location and rotation data. Figure 5.8 shows how the raw input data is converted into a high-dimensional feature vector.

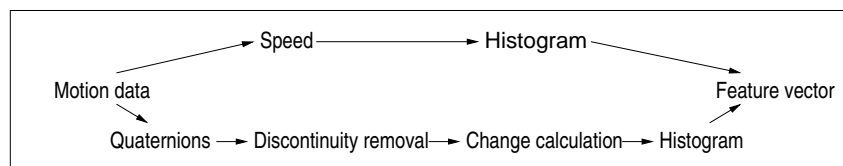


Figure 5.8: The current motion preprocessing.

An artificial neural network should transform the feature vector into estimation of the conductors emotion. To do so we need to describe the emotions in a computational way. From prior research we identified three mod-

els that could be used. First approach is to assume that each emotion is one dimension in high-dimensional space (e.g. [19]). The second is to assume that emotions are mutually exclusive (e.g. [37, 56, 129]). In this model a human has one active emotion. The third model uses a low-dimensional emotional space to model emotions (e.g. [118, 61, 75, 134, 166]). An example of such model is in Figure 5.9.

In P4 all three models are used. There were two conductors whose motion data was analyzed with ANNs. For one conductor the analysis system could identify the gestures fairly reliably. For the other the system had much more trouble. As an example Figure 5.10 shows how the system mapped different emotions for the easier conductor.

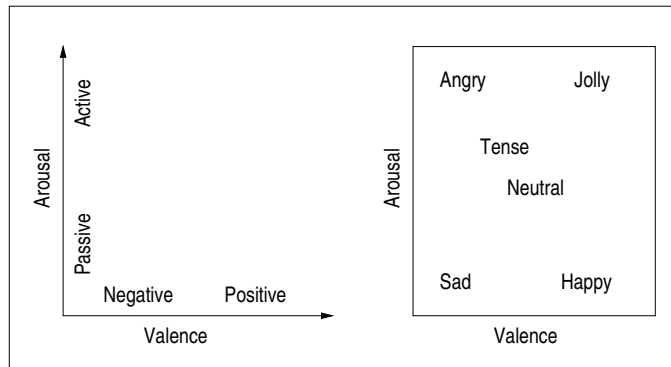


Figure 5.9: Definition of arousal/valence space and positions of emotions used in the space.

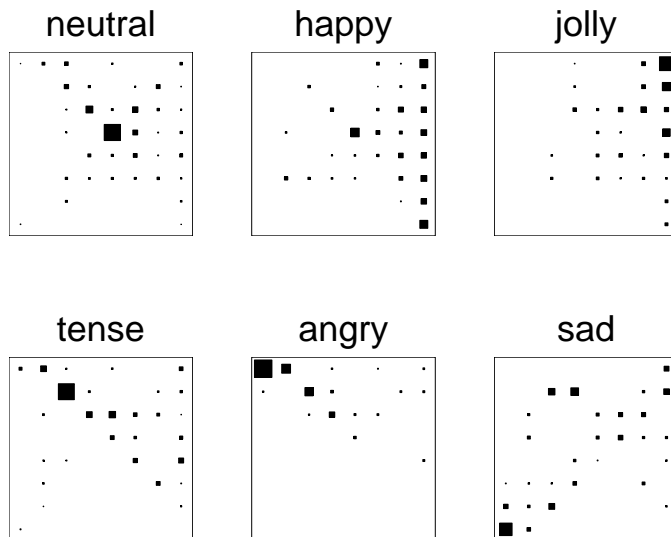


Figure 5.10: Distribution of ANN-estimations in arousal/valence space. These should match the locations of these emotions in figure 5.9.

Player Modeling

Recognizing tempo and controlling the music with this tempo has been the goal of the music synthesis. The tempo should change swiftly when conductor wants to change it, but it should stay stable in other cases, even if the movements of the conductor don't match an exactly even tempo. Fulfilling these contradicting aims at the same time is one of the main problems of the player modeling.

The tempo control method is implemented as a state machine. The player model has an internal state that consists of current score position and tempo. The system has a number of input parameters: the expected and estimated tempo, expected and estimated nuances, the score and the desired expression style.

The base tempo estimation is taken from beats - the time and score difference between down-beats indicates the tempo. Finer adjustments are done constantly depending on the estimated beat phase. If the follower appears to be lagging behind (or advancing too much) both tempo and score position are adjusted to counter the effect. The strength of the adjustment can be varied depending on the score, tempo, conductor's gestures and musical hints.

When many notes are played quickly in succession tempo changes become more audible. As an opposite, if a long pause is about to end the first note-on event can be played without concern about the past tempo. Player sensitivity to tempo changes is automatically decreased as the note density increases to compensate for this.

The playback volume affects the MIDI note velocity of the notes. Increased or decreased volume simply causes an offset to the MIDI velocity values. The player model can also produce staccato effects. When the conductor is playing in staccato style all short notes (up to a quarter note) are made even shorter.

DIVA System in Use

The DIVA conductor follower has been presented twice in large scale. The first time was in SIGGRAPH'97 conference where it was one of the main installations in the Electronic garden. This was a big and heavy deployment with magnetic motion tracking for conductor following, graphics on a large screen, physically modeled instruments and extensive 3D audio processing.

Over the one-week period the DIVA stand was often full with enthusiastic people. Many TV channels included it in their news shots, displaying the entertainment value of the system.

The second presentation of the conductor follower was in the Heureka science center in Finland. We tried to make the installation as robust as possible by making a new motion tracking mechanism that could be used unattended. In practice children still damaged the baton and the system required frequent inspection and repair. When DIVA was working it was clearly the most popular installation in Heureka.

Installations like these allow ordinary people — children and grown-ups — to use the system and play the role of a conductor for moment. The shortcomings of the system in interpreting the intentions of the conductors are not serious in these settings, since the general idea is to have fun, not

create an artistically convincing performance.

In spite of the original intentions the virtual orchestra has not been used as a tool in conductor training. It has been tested in concert settings for a few times, but these settings reveal and emphasize the problems that do not disturb the system when used in pure entertainment context.

5.2 Helma

Helma was a project to develop an immersive 3D drawing/sculpting application. AnimaLand software served as the technical starting point — Helma began as a series of extensions to the AnimaLand. Another important source for inspirations was the Antrum application. Antrum was made for Wille Mäkelä by a group of students in HUT. It was an application that could be used to draw polygon models [95].

The making of Helma has resulted in several publications, some of which are part of this thesis. Publication P5 describes the use of the virtual pocket user interface metaphor (Section 5.2). Helma has also resulted in other publications that cover the project's artistic [95, 96] and technical aspects [127]. Examples of art made with the Helma system are in Figures 5.11 and 5.12.

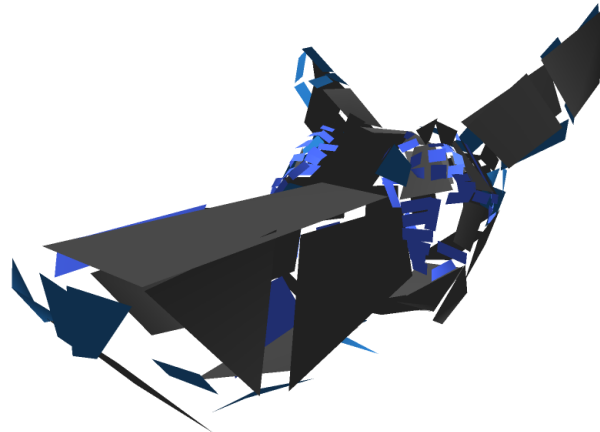


Figure 5.11: A dog sculpture by Joakim Sederholm.

One of the central aims of the Helma project was to explore new interaction methods. The application has two main modes for the user interaction. Drawing mode is used for crafting the graphics directly. This mode offers direct-manipulation approach to making graphics. Control mode is used to select and configure the tools (sprays, magnets, meshes, etc.). Originally the main instrument of this mode was the “kiosk” that contained controls for all of the drawing tools (Figure 5.13).

While the kiosk is a powerful control center, it is also a source of problems. Its use interrupts the drawing process and it may appear in an awkward position. The virtual pocket metaphor was seen as a way to overcome these problems.

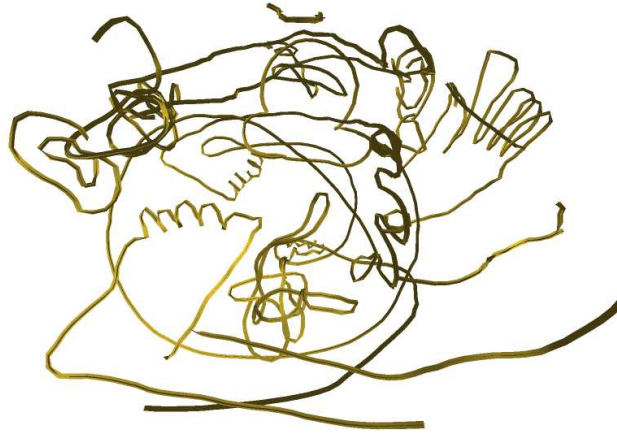


Figure 5.12: A portrait by Jukka Korkeila.



Figure 5.13: A user with the kiosk.

Virtual Pockets

Most computer applications have a simple basic logic: The user has a collection of tools that are used to achieve some goal. These tools should be easy to use and easy to select. The latter one is often a problem in VR applications. Publication P5 describes how virtual pockets can be used for tool selection in VR.

Virtual pockets mimic physical pockets. They have a location that is directly related to the user's body. The user can take tools from the pockets and put new tools into them. Unlike physical pockets one can store only one tool into one virtual pocket. The application may use a feedback mechanism for letting the user know when he/she hits a pocket and what tool is in the pocket.

We developed our virtual pocket implementation for the Helma system.



Figure 5.14: Visual area display in an otherwise empty world. The first pocket is on top of the left shoulder and fifth on top of the right shoulder (for right-handed users).

In Helma tool selection is often followed by tool configuration. To support this work-flow the tool configuration was integrated into the virtual pockets: In addition to tool selection the user can also adjust the most important parameters of the tool.

We put the pockets around the user's head, as shown in Figure 5.14. Using head as the origin was motivated by the desire to minimize the number of motion tracking sensors (head already had a sensor) and to make the system work without calibration for people of different sizes. Another important factor was that people seldom put their hands close to head while drawing. As a result pocket misactivations are rare.

Without any feedback most people find the pockets difficult to use. To test different feedback methods we ran usability tests. In the tests people had five feedback options: no response, area display (shown in Figure 5.14), area display with tool names, audio feedback and area display with audio feedback. Based on both user preferences and objective measurements it seems that visual feedback with textual information is the most useful feedback method. There were strong individual differences, which implies that the system response should be configurable to match different users.

Upponurkka

While the previous systems were tools for the creation of art, Upponurkka is an installation that is ready for the audience. Upponurkka installation was displayed in Kiasma — Finland's premier modern art museum — in November 2005, and in the Ars Electronica festival in Linz, Austria in 2006. The aim was to create a cost-effective system that would allow ordinary people to experience the 3D graphics with a user interface that resembles the one that was used to create the graphics.

Overview of the Upponurkka system is in Figure 5.15. The hardware components are listed below:

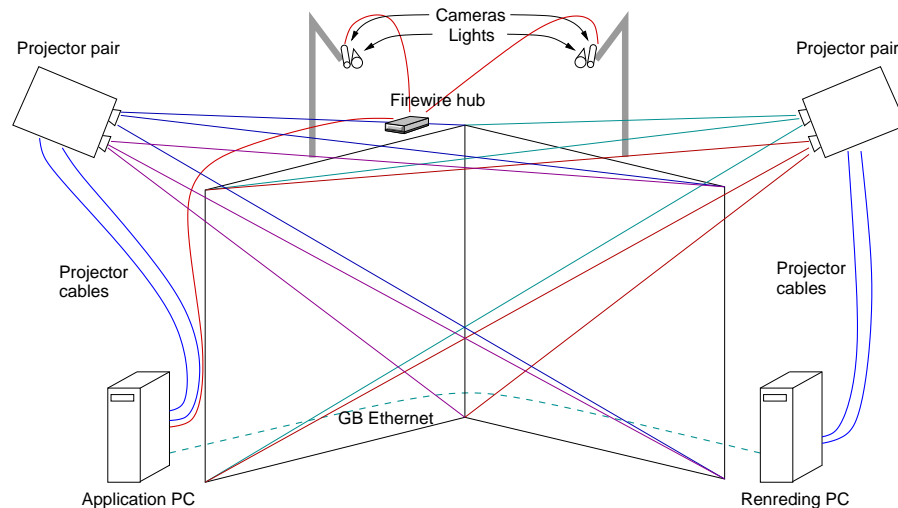


Figure 5.15: Overview of the components used in the Upponurkka system

- The most visible part of the system are the two 3x3m silver screens.
- The image to these screens comes from two pairs of projectors. The projectors have polarizing filters in front of their lenses. Currently the projectors are common LCD projectors. LCD emits polarized light. The polarization axis should be tilted 45 degrees so one can change the projector polarization into vertical or horizontal with a simple filter.
- The image to these projectors comes from two computers — the application PC and the rendering PC. Each computer has a single graphics card with dual outputs and both are responsible for rendering 3D content to one pair of projectors. The PCs have integrated GB ethernet controllers that takes care of their communication.
- There are two Firewire color VGA cameras that are used for motion tracking.
- There are lights next to the cameras (as close as possible)
- Cameras are connected to the application PC via a Firewire hub.
- Polarizing stereo-glasses are needed to view the graphics properly. The glasses that are to be tracked should be equipped with balls that are covered with retro-reflective material (e.g. stuff from the Taperoll company).

The camera-based motion tracking was implemented as a Fluid-plugin after which we could leverage the existing Helma infrastructure. The graphics were distributed between the two computers using Broadcast GL.

A visitor with the tracked glasses could select a set of works to view. There were three sets with different works and the works were changed according to a predefined schedule.

From the artistic perspective the installation gave mixed results. A vast number of the visitors were impressed by the head-tracked 3D graphics,

which were a novelty for them. Unfortunately far smaller number of people became interested in the actual art with only a small fraction of people looking at all (or most) available works. The people who were interested in the works typically moved about, seeking new angles to view the works. This supports the original assumption that an immersive large-scale display would allow people to use the system in more bodily way.

A more detailed description of the system and some analysis of user behavior has been published by Lokki et al [93].

5.3 Kylä

Like *upponurkka* *Kylä* is in art installation for the general audience. *Kylä* can be viewed as art or educational system that presents ancient folk art to modern audience. The interaction and staging are used to elicit an atmosphere that supports the rune-singing.

Technically *Kylä* is a straightforward installation. The motion of the single candle is tracked indirectly with two or more cameras. The cameras are located as high as possible and they are directed to the pictures over the visitors' heads. The candle tracking code calculates the average luminance values in each of the pictures. The luminance values are averaged over a few seconds to counter the problems of camera noise. Since different images have different brightness the luminance values are scaled by image-specific normalization coefficient. The candle is assumed to be closest to the picture with the highest luminance.

Once the application knows where the candle is it needs to play the sound-track that is related to the relevant picture. The audio player is implemented as a *Mustajuuri* plugin. This enabled us to use the ordinary audio plugins to enhance the signal with equalization (to overcome camping caused by the dark cloth) and reverberation (to make the soundtracks slightly more spacious).

From HCI-perspective *Kylä* is a novel system since it uses fire as a user interface element (see Figure 2.6). Candles are seldom used as interaction devices and even when used, they tend to be stationary part of the installation, as is the case with the work by Hlubinka et al. and Burleson et al. [60, 24]. *Candlelight* has many interesting properties which contribute to the installation. Firstly a small flame alone creates a peaceful atmosphere that is necessary for the appreciation of the content of the installation. Secondly it sets limitations for movement, visitors automatically move slowly when a candle is present. Thirdly it is part of a movement-based user interface. The visitors are in control, but they need to be active to change the song. Finally it makes the installation more social. Since there is only one candle visitors tend to gather around one picture at a time.

From a visitors perspective *Kylä* is a room that has something to sing and to show. The technology is invisible. It can be seen as a case of ambient [114], ubiquitous [160] and embodied [38] interaction. It is also a case of holistic design. If one of the elements was missing *Kylä* would not be sensible any more.

The two deployments (media art festival in 2005 and open-air museum in 2006) provided feedback about how well the design goals were met. In

the media art festival Kylä was well received and it had much of the effect we had planned. When visitors heard the child's voice visitors would smile, while the picture with graveyard resulted in solemn expressions. Majority of the people liked the system. The design proved to be surprisingly inclusive, with people praising the atmosphere of the installation regardless of their age, gender or nationality.

In the open-air museum the installation was in the right cultural context which supported stronger experiences. Roughly 5–10 people left the installation either crying or holding back tears. It seems that the experience was positive for most (possibly all) of these people. This behavior was most common with Finnish women (particularly older women), but others reacted strongly as well (younger people, some men and foreigners). That the installation communicates well to older people is an interesting finding. These are the people who are seldom reached by modern high-tech systems. While the installation was in the museum the word about it began to spread and many visitors came to the museum just to see Kylä, again showing that its popularity.

Observation and discussions with visitors indicate that the strong emotional involvement is related the careful combination of several elements that contribute to the total interaction design. In the installation the physical artifacts (candle, pictures), songs and visitors' bodily behavior work together to create a strong experience. The subtle use of computer technology enabled us to build a novel fire-based interaction system where the archaic culture is presented in a way that speaks also to the modern visitor.

The content of the installation is strongly localized to a particular place (Viena Carelia) and time (19th century). In spite of this the experience was shared in a meaningful way across several barriers: age, gender and nationality.

5.4 Discussion

This chapter described four applications for artistic, embodied, multi-modal interaction. These applications have very different roles in the artistic process:

The conductor follower is a system that simulates natural human-to-human interaction. In this work the emphasis has been on detecting features that conductors use in their ordinary work and react to them as a real musician would. The system has been demonstrated in live performances. We have also made a low-cost entertainment versio where children can use it to conduct music. Compared to other conductor following systems our approach introduced several new ideas: artificial neural networks, more advanced time signature handling and recognition of nuances and emotional cues.

Helma is another artictic application, but this time an independent system that is used to create and display 3D graphics. The user interface does not mimic any natural interaction, but follows rules of its own. Due to this we have developed a comprehensive VR user interface and also done user testing on some of the user interface ideas. Its predecessors are other virtual reality 3D painting applications, in particular Keefe's [77] and Schkolne's

[135] systems. In more wide sense also the desktop 3D applications can be seen as belonging to the same group. Compared to its predecessors and contemporaries Helma offers a greater variety of tools and algorithms, most of which have been tuned especially for the immersive drawing purposes. It also includes animation features which have been lacking from other VR applications.

The Upponurkka installation is a system to show the art that is made with the Helma system. Technically the installation worked as planned, but as art it did not fulfill our goals.

Kylä is an independent art-work that with subtle physical interaction and audio feedback. Its predecessors are installations that have presented similar content in a museum-like, static environment. In this thesis Kylä is a support case that verifies that the technologies we created are useful in different applications. From artistic perspective this work was more powerful than expected.

Side-by-side comparison of Kylä and Upponurkka reveal interesting differences and similarities. Both are simple installations where the visitor's only interaction method is to simply move around, requiring some bodily activity from the visitors. Both aim to present media that can be difficult to communicate in an effective way. In the case of Kylä it is archaic unaccompanied rune-song that even in Finland is foreign to most people. The 3D art that is presented in Upponurkka is likewise difficult to present to ordinary people without the aid of stereo graphics. The intended atmosphere of the installations and the role of technology in them is very different. Upponurkka is a technologically sophisticated installation that displays modern art made with futuristic tools. Kylä in turn reaches to the past with minimal technological overhead.

The experiences in the exhibitions show that Kylä reached closer to the goals that we had set for it. Above are listed some differences that might be the cause of the differences. However, without thorough studies on the visitor experience it is not possible to evaluate why exactly Kylä and Upponurkka elicited such different responses.

Together these four applications present different approaches on multimodal interaction. They exemplify the strengths and weaknesses of embodied interaction in different contexts.

6 TECHNIQUES FOR AUDIOVISUAL EFFECTS

All interaction concepts in this thesis require audio and/or visual effects. The effects can serve as feedback to the user or as special effects to enhance the environment. To create the necessary effects new techniques were developed.

6.1 Mustajuuri Audio Engine

In many of the visions sound is an important medium. Since most of the systems were designed to work in a VR environment we needed a tool that would be able to create the relevant audio effects in real time. The system needs to perform multiple tasks:

1. Synthesize sounds in real-time. The sound synthesis can range from simple sample playback to more complex synthesis systems (e.g. granular synthesis).
2. Locate the sound to arbitrary 3D position with VBAP, distance delays and distance attenuation.
3. Allow smooth interpolation of all parameters.
4. Apply individual loudspeaker equalization. Since the loudspeakers are behind the walls in the EVE their frequency response needs to be corrected with proper filtering.
5. Apply individual loudspeaker gain correction. The volume of the loudspeakers must be calibrated. This task is helped by in-application gain correction routines.
6. Work with low latency.

To meet these needs Mustajuuri was developed. Its basic architecture is described in publication P6. An overview of how Mustajuuri fits our auralization scheme is described in Hiipakka's work [59, 58]. Mustajuuri has also found its way to other VR environments where it is used as a 3D audio engine [79].

Mustajuuri is basically a plugin-based audio processing engine. This is a generic architecture that can be used in VR systems (AnimaLand, Helma), music mixing or as a custom sound engine (Kylä). Its novelty lies in the powerful plugin framework. Each plugin (or module) has two kinds of input and output: audio signals and control messages (Figure 6.1). Audio signals are passed as floating point buffers from a plugin to another. The control messages are delivered when one needs to change some internal parameter of a plugin. The messages are timed with sample-accurate timestamp. This time-stamping allows smooth and accurate interpolation of parameters. The control messages contain two parts: a character string (parameter name) and message object (arbitrary data). The control framework is influenced heavily by the Open Sound Control (OSC) communication protocol [164].

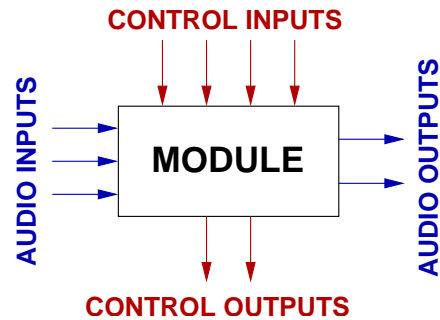


Figure 6.1: The inputs and outputs of a Mustajuuri module.

In Mustajuuri the plugins have more power than they have in other systems (e.g. Pd or commercial audio applications). For example plugins can access other plugins in the signal processing graph and create links that are not part of the normal signal/message transfer path.

Mustajuuri has a GUI that is used to configure the system. While GUI is not strictly needed for VR systems it has proven to be very useful since it can be used for rapid adjustments.

Discussion

Mustajuuri shares many features with preceding audio signal processing platforms: It can be extended dynamically with plugins and it offers a GUI for signal routing and parameter adjustment.

The differentiating feature is the comprehensive plugin system. The novel features of this approach are:

- In Mustajuuri plugins can access other plugins in the DSP graph. This feature is useful when making applications that require tight cooperation of several custom plugins. For example an auralization control plugin uses direct links to sound synthesis and auralization control plugins.
- Plugins can affect the main application, by e.g. starting and stopping the DSP engine. This allows one to turn plugins into custom applications that utilize the Mustajuuri framework. For example the Kylä software is implemented as a custom plugin that controls the application.
- The number of a plugin's signal channels and available parameters is dynamic. This allows one to make plugins that offer a dynamic set of parameters — for example an HRTF auralizer plugin that can be run with very different parameters.

Together these features allow one to use Mustajuuri in flexibly in different ways. It also offers a plugin developer different levels of interaction and control when writing plugins.

6.2 Graphics Techniques

Second Order Particle Systems

The power of particle systems lies on their ability to use a large number of simple particles to mimic complex natural phenomena. The simple particles are guided by forces that affect their behavior. As an example Figure 6.2 shows a fire effect that was made with traditional particle systems. This example also reveals one of the problems of particle systems. If the forces are static then the final animation looks artificially static.

So far the solutions to this problem have been to animate the forces manually or use a video billboard with more complex animation. The first solution is problematic since it requires manual work and the second because videos lack volumetric feel that is required from particle systems. Neither approach works if one wants to make animations that react dynamically to the environment.

These problems can be avoided by making the forces dynamic. Publication P7 describes an automatic approach for this task — the second order particle system. In this approach all forces are attached to dedicated force objects (or particles). The objects share the physics of the particles. This leads into more dynamic animation as the forces move about the scene. Figure 6.3 shows how moving vortices affect the fire animation of Figure 6.2. In this case the animator needs to define the behavior of the particles (mass, color changes etc.) and the behavior of the forces (the type of force objects and their parameters). Manual control of the moving vortices in this fire effect would be a heavy task.

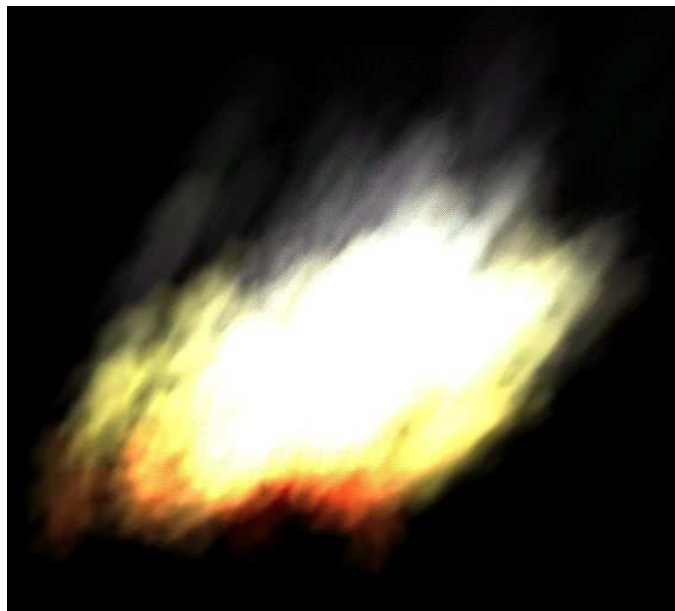


Figure 6.2: Fire effect with a traditional particle system.

A typical way to use the second order particle system is to create force particles into the world with slightly randomized parameters. This randomization guarantees that the effect does not repeat any animation sequence.



Figure 6.3: Fire effect with the second order particle system.

Broadcast GL

Some applications need to run several graphics displays at once. For example in a CAVE-like virtual reality systems the same virtual world needs to be rendered multiple times from different viewing angles. There are two ways to generate the graphics in these cases. One is to use a computer with a sufficient number of graphics adapters. The second one is to use several computers that together render the graphics. The latter is expected to provide better price/performance ratio, but it requires clustered rendering. Publication P8 describes how to achieve efficient clustering with off-the-shelf components.

This approach is efficient if the following assumptions hold:

1. The displays share significant part of their 3D content.
2. The capacity of the clustering network is a critical factor in the graphics distribution. In other words the graphics distribution stresses the network resources.
3. The application developer wants to keep the application monolithic and use the normal graphics programming interfaces (e.g. OpenGL).

Based on these assumptions we developed Broadcast GL (BGL). This is an OpenGL distribution system that uses UDP broadcast to transmit rendering commands from the application to the rendering computers. As a return channel from the renderers BGL uses TCP sockets. OpenGL API was chosen because we already had it in use and because it has been designed to be easily streamable over networks. An overview of the system is in Figure 6.4.

Benchmarks prove that BGL offers high scalability. Its performance drops minimally as the number the number of rendering computers is in-

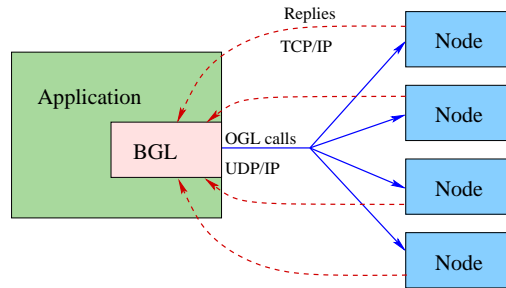


Figure 6.4: The OpenGL distribution mechanism used in BGL.

created. In many cases it also utilizes network resource more efficiently than the preceding systems.

Discussion

The second order particle system is as an extension of the traditional particle systems. Earlier dynamic particle effects have been created by moving the forces fields manually in the scene. This approach is cumbersome if one needs to create longer animation sequences. The dynamic forces are a powerful tool for building animations that are dynamic, but do not repeat themselves.

Our particle engine was originally developed to be used in the Storm installation. With this system a user could create different atmospheric effects with gestures. The system would map gestures to control parameters that would control a second order particle system. In the AnimaLand application the particle system was in use, but the more advanced features (independently moving forces etc.) were not put to use due to time constraints.

Broadcast GL is another graphics system that was created during the thesis work. The concept was originally invented to be used in the EVE virtual room in HUT/TML, but the first real deployment was done in the Upponurkka installation. The main predecessors of BGL are the WireGL and Chromium systems. Compared to them BGL adopts a more simple UDP-broadcast approach to graphics distribution. Bechmarks against the Chromium support our original assumption that such simple method works equally well or even better for many applications.

The presented techniques and concepts could be pursued further in many ways.

Multimodal input management systems could be developed further to make the systems easier to develop. Based on the experience gained with FLUID it seems that the most effective way to support input device and data management is to create simple, uniform building blocks that the software developer can chain together in the software. Such building blocks could be input device drivers, signal transformers and pattern recognizers. While FLUID includes many of these ideas they could be pushed further by providing each component in an isolated module.

Different conductor following systems have been built around the world. The level complexity in the systems has increased and people have also made conceptual contributions to the field, such as calling for more holistic reasoning in the conductor following engine. Right now the weakest point of the development seems to be the lack of real-world evaluation of the technologies. After all, a complex state-of-the-art system does not necessarily work better than a simple hack. Such testing, possibly in the form of a competition, would direct the research and guarantee that the new approaches work not only on paper, but also in the field.

The emotional information that can be extracted from the conductor motion is an interesting side-track of the conductor follower. Obviously the information should be used in the music synthesis, giving the system more artistic potential. Of course, the above criticism that the systems should be tested with some real users and applications still holds.

The use of accelerometers to track conductor motion could be refined, but it does not seem like a promising source of new scientific findings. Instead one could devise either new tracking methods or apply the accelerometer-based approach to some other domains where slightly inexact motion tracking is sufficient.

Immersive 3D graphics creation systems are an interesting topic. Unfortunately the required hardware and software infrastructure is so heavy that it makes progress quite difficult — a few projects, scattered around the earth cannot really do continuous research around the topic. A more promising future for 3D graphics interaction would be to seek application concepts in the field of entertainment, for example in science centers and amusement parks, where it is possible to build large- or medium scale installations. Here one could use the Upponurkka installation as a starting point, possibly augmented by tangible interaction.

The virtual pocket metaphor could be adapted to a wider variety of virtual reality applications, where selection tasks are common. The approach could also be useful in combination with smart clothing.

Mustajuuri has demonstrated how to build an integrated framework for interactive audio processing. The development of such frameworks is in itself not very interesting, but rather the use of such systems to aid rapid application development. Like FLUID the problem is that different applica-

tions have different, contradictory needs. Possibly the best way to overcome this problem is to provide a platform that is open enough to be extended to meet the requirements of a particular application. In this respect the current systems provide different compromises.

Of the technical innovations in this thesis the second order particle system is possibly the system with greatest use potential. It could be developed further by 1) creating new force particles, 2) optimizing the behavior of the current forces and 3) optimizing the rendering process, and 4) combining 2nd order particle system with physical modeling. Some of this work has already been done by the author in later publications on particle system rendering [67] and dynamics calculation [66].

The Broadcast GL system could be developed to include the latest OpenGL features (OpenGL shading language, vertex buffer objects etc.) to the extent that it is possible. In a more fundamental scale it could be developed to offer even better scalability by the use of broadcast chaining via proxies.

The concepts could also be developed further or one could create variations on them.

The DIVA system with the conductor following could probably be made (even) more fun for the general audience. The versions that were displayed in SIGGRAPH and Heureka were built without any real user testing. A more thorough evaluation of these systems would probably reveal new ways to improve them. Another alternative would be to move the applications into some other areas, for example making a webcam-based version that would run on a normal PC.

The Helma system would always benefit from more 3D tools, although there is risk that the system becomes too complex to use. More productive approach might be to (again) perform tests with more users and possibly involve also more user interface specialists in the design process.

The Upponurkka system worked technically as planned, but the user experience was not as strong as one would hope. One could try to use the existing hardware and software infrastructure to create different applications, for example games.

There is no obvious way to improve the Kylä installation. Observations in the field show that it was well received and any changes are likely to be for the worse. The best future direction would then be to use the same template to present other cultures, not only the archaic Carelia. The candle is a powerful interaction device, but other alternatives to select the media could also be tested, for example touching.

This thesis presented several novel multimodal concepts and their technology. As these concepts were planned and implemented we needed to create several new support tools and technologies for these applications: user input management, gesture analysis, audio processing and computer graphics algorithms. In many cases the new systems were improved versions of existing techniques and architectures.

These new techniques were needed for the development of novel multimodal applications. As such they demonstrate how the basic technological components become intertwined in this kind of applications. Often one is forced to develop some new technology to overcome a barrier in the concept implementation process. This thesis reports cases where this development has been successful, resulting in systems that can be put public display for weeks or months, often with minimal maintenance.

The main effort has been on technology development, but implementation of the concepts has involved multi-disciplinary work as well. In the Kylä installation this took place as concept and content planning, in the Upponurkka installation concept planning and in the the Helma system as user interface design. In the scope of the projects there have been people with different backgrounds, leading into even more disciplines being present, from 3D modeling to 3D art and animation.

The application of these techniques in the conductor follower, Helma, Upponurkka and Kylä concepts has given results that can be reflected against the initial motivations of this thesis. All of these concepts include gestural, bodily interaction in one way or another.

Helma was developed to investigate the possibilities and limitations of combining bodily interaction with 3D graphics. The system has been used primarily by Wille Mäkelä, with six other visiting artists.

The Upponurkka installation provides the general public with a user interface to experience 3D art — either animations or static sculptures.

In the context of public installation the conductor follower enables ordinary people to act as conductors. Typically the users come as groups, leading to an enactive social event.

Compared to the conductor follower Kylä is at the opposite end of the spectrum. Instead of large movements the visitors are encouraged to move slowly and carefully around a space that is augmented with interactive camera-based sound rendering.

Not all of the systems can be counted as completely successful, but the successful elements in them display the power of multimodal interaction. Physically active interactive systems promote and exemplify the idea that physical and mental processes are coupled. These systems have demonstrated the potential to create different kinds of experiences to the users with the aid of multimodal interaction. The types of experience range from having fun socially (conductor follower), immersive 3D art creation (Helma system), 3D interactive art experiences (Upponurkka) to tranquil exploration of an archaic culture (Kylä).

For each of these experiences it is difficult to find a direct match from the more traditional desktop computing. This supports the original motivation of this thesis: The assumption that multimodal interaction can provide new experiences that cannot be supported with traditional mouse-and-keyboard interfaces. The concepts are based implicitly on the idea that mental and physical activity are directly related.

Of the four systems Kylä stands out as being different, even paradoxical, in its way to use technology. It is a high-tech installation where the technology is hidden. Instead of predicting or problematizing modern trends Kylä reaches back through millennia. It is an attempt to ask (and possibly answer) in a profound way who we are, and what is our heritage. If Kylä and other systems have given some people joy, then this work has caused not only scientific and technical progress, but also something good.

9 MAIN RESULTS OF THE THESIS AND CONTRIBUTION OF THE AUTHOR

Publication [P1]

Software Architecture for Multimodal User Input – FLUID. The creation of multimodal interaction has been and remains difficult. One part of the problem is that handling user input data is very difficult. There are numerous systems that support various input devices, but they typically do not maintain data quality. In most applications it is necessary to process the data as equally sampled series.

To handle these needs we created a new architecture and an implementation of this architecture – FLUID. This architecture is composed of an input layer that manages the input hardware and a data-processing layer that is used to analyze the data with modular approach. This system was needed for the creation of the AnimaLand environment. It also raises discussion about basic issues of human-computer interaction: What is input and how does the human input turn into an action within the machine.

The author designed the basic architecture and implemented the data processing layer while Janne Kontkanen designed and implemented the data-input layer.

Publication [P2]

Conductor Following With Artificial Neural Networks. This paper represents a conductor following system that analyzes the conductor’s hand motion and plays music accordingly. The most significant idea of this research is the combination of heuristics and artificial neural networks.

The author created, implemented and tested all algorithms in this work.

Publication [P3]

Accelerometer-Based Motion Tracking for Orchestra Conductor Following. This paper describes a method for using cheap accelerometer-based motion tracking in the context of conductor following. Typically accelerometers need frequent calibration to counter the inevitable drift caused by the technique.

This paper presents how we can overcome this problem by using case-specific signal processing. The usefulness of the signal processing algorithms depends on the way they take into account the known properties of the conductor motion.

The author designed and implemented the signal processing algorithms, while Janne Jalkanen was responsible for the hardware design.

Publication [P4]

Detecting Emotional Content from the Motion of an Orchestra Conductor. Emotionally sensitive — or affective — software could improve the quality of the human-computer interaction. The very first step towards that direction is detecting the emotional state of the user.

In this paper we present methods for analysis of the emotional content of human movement. We have studied orchestra conductor's movements that portrayed different emotional states. Using signal processing tools and artificial neural networks we were able to determine the emotional state intended by the conductor. Multiple definitions of the emotional space are tested for relevance and performance: Mutually exclusive emotions, orthogonal space and low-dimensional space.

The author invented and implemented the methods used in this paper and ran the user tests.

Publication [P5]

Virtual Pockets in Virtual Reality. This paper presents a new user interface metaphor that is especially suited to virtual reality applications. It is expected to generalize into other domains as well: mobile devices and tangible interaction.

The author implemented the virtual pockets. The behavior of the pockets and the user tests were done with Markku Reunanen.

Publication [P6]

Mustajuuri - An Application and Toolkit for Interactive Audio processing. This paper presents the sound processing engine that was developed to handle the audio processing in the virtual room. While that is one its uses it has been designed to be general enough to be useful for almost any audio processing task.

Mustajuuri was made completely by the author.

Publication [P7]

The Second Order Particle System. This paper presents an extension to the classical particle systems – the second order particle system. The most important new idea in this paper is the inclusion of interactive forces. This means that forces affect each other, resulting in much more lively and credible animations.

The author is responsible for most of the paper. Janne Kontkanen proposed the idea of creating burning material and contributed some source code for it.

Publication [P8]

Broadcast GL: An Alternative Method for Distributing OpenGL API Calls to Multiple Rendering Slaves. In virtual reality systems it is often necessary to render the same data sets from different angles. The traditional method to handle the task is to use a single powerful computer. A much cheaper method is to use a cluster of off-the-shelf PCs that are tightly synchronized.

Our approach is to distribute the graphics API calls via a UDP/IP broadcast or multicast. The use of network broadcast makes this approach highly scalable. We present overview of the system and show its performance against other systems.

The author implemented the system and also largely invented the methods that are used to achieve high scalability.

BIBLIOGRAPHY

- [1] Jérémie Allard, Valérie Gouranton, G. Lamarque, Emmanuel Melin, and Bruno Raffin. Softgenlock: Active Stereo and Genlock for PC Cluster. In *Proceedings of the Joint IPT/EGVE'03 Workshop*, Zurich, Switzerland, May 2003.
- [2] Jérémie Allard, Valérie Gouranton, Loïck Lecointre, Emmanuel Melin, and Bruno Raffin. Net Juggler: Running VR Juggler with Multiple Displays on a Commodity Component Cluster. In *VR '02: Proceedings of the IEEE Virtual Reality Conference 2002*, page 273, Washington, DC, USA, 2002. IEEE Computer Society.
- [3] Jérémie Allard, Bruno Raffin, and Florence Zara. Coupling Parallel Simulation and Multi-display Visualization on a PC Cluster. In *Euro-par 2003*, Klagenfurt, Austria, August 2003.
- [4] Antonella De Angeli, Walter Gerbino, Giulia Cassano, and Daniela Petrelli. Visual Display, Pointing, and Natural Language: The Power Of Multimodal Interaction. In *AVI '98: Proceedings of the working conference on Advanced visual interfaces*, pages 164–173, New York, NY, USA, 1998. ACM Press.
- [5] Ascension. Ascension Technology Corporation, 1999. Company's WWW home page, URL: <http://www.ascension-tech.com/>.
- [6] Themis Balomenos, Amaryllis Raouzaïou, Kostas Karpouzis, Stefanos Kollias, and Roddy Cowie. An Introduction to Emotionally Rich Man-Machine Intelligent Systems. In *Proceedings of the 3rd European Symposium on Intelligent Technologies, Hybrid Systems and their Implementation on Smart Adaptive Systems (EUNITE'03)*, Oulu, Finland, 2003.
- [7] Allen Bierbaum, Christopher Just, Patrick Hartling, Kevin Meinert, Albert Baker, and Carolina Cruz-Neira. VR Juggler: A Virtual Platform for Virtual Reality Application Development. In *The Proceedings of IEEE VR Conference 2001*, 2001.
- [8] Mark Billinghurst. Put that where? voice and gesture at the graphics interface. *SIGGRAPH Comput. Graph.*, 32(4):60–63, 1998.
- [9] Oliver Bimber, L. Miguel Encarnaç o, and Andr e Stork. A multi-layered architecture for sketch-based interaction within virtual environments. *Computers & Graphics*, 24:851–867, 2000.
- [10] Richard A. Bolt. "Put-that-there": Voice and Gesture at the Graphics Interface. In *SIGGRAPH '80: Proceedings of the 7th annual conference on Computer graphics and interactive techniques*, pages 262–270, New York, NY, USA, 1980. ACM Press.
- [11] Richard Boulanger, Young Choi, and Luigi Castelli. The Mathews' Radio-Baton: A Windows & Macintosh GUI. In *Proceedings of the 2000 International Computer Music Conference*, 2000.
- [12] Richard Boulanger and Max Mathews. The 1997 Mathews Radio-Baton & Improvisation Modes. In *Proceedings of the 1997 International Computer Music Conference*, pages 37–46, Thessaloniki, Greece, 1997.
- [13] Doug Bowman, Joseph Gabbard, and Deborah Hix. A Survey of Usability Evaluation in Virtual Environments: Classification and Comparison of Methods. *Presence: Teleoperators and Virtual Environments*, 11(4), 2002.

- [14] Doug A. Bowman and Larry F. Hodges. An Evaluation of Techniques for Grabbing and Manipulating Remote Objects in Immersive Virtual Environments. In *SI3D '97: Proceedings of the 1997 symposium on Interactive 3D graphics*, pages 35–38, New York, NY, USA, 1997. ACM Press.
- [15] Doug A. Bowman, Donald B. Johnson, and Larry F. Hodges. Testbed Evaluation of Virtual Environment Interaction Techniques. *Presence: Teleoperators and Virtual Environments*, 10(1):75–95, 2001.
- [16] Doug A. Bowman, David Koller, and Larry F. Hodges. Travel in Immersive Virtual Environments: An Evaluation of Viewpoint Motion Control Techniques. In *VRAIS '97: Proceedings of the 1997 Virtual Reality Annual International Symposium (VRAIS '97)*, page 45, Washington, DC, USA, 1997. IEEE Computer Society.
- [17] Marshall Brain. *Motif Programming: The Essentials... and More*. Digital Press, 1992.
- [18] Bennet Brecht and Guy Garnett. Conductor Follower. In *Proceedings of the International Computer Music Conference*, pages 185–186, San Francisco, 1995. Computer Music Association.
- [19] Roberto Bresin and Anders Friberg. Emotional Expression in Music Performance: Synthesis and Decoding. In *KTH, Speech, Music and Hearing, Quarterly Report*, volume 4, pages 85–94. Kungliga Tekniska Högskolan - Royal institute of technology, Stockholm, Sweden, 1998.
- [20] Roberto Bresin and Anders Friberg. Emotional Expression in Music Performance: Synthesis and Decoding. *Computer Music Journal*, 24(4):44–62, 2000.
- [21] Kenneth R. Britting. *Inertial Navigation Systems Analysis*. Wiley & Sons, Inc., New York, 1971.
- [22] Frederik P. Brooks. What's Real about Virtual Reality? *Computer Graphics and Applications*, 19(6):16–27, 1999.
- [23] Grigore Burdea and Philippe Coiffet. *Virtual Reality Technology*. John Wiley & Sons, 2003.
- [24] Winslow Burleson, Paul Nemirovsky, and Dan Overholt. Hydrogen Wishes. In *SIGGRAPH '03: Proceedings of the SIGGRAPH 2003 conference on Sketches & applications*, page 1, New York, NY, USA, 2003. ACM Press.
- [25] CAVELib. CAVELib User's Manual. WWW-page, Cited 24.10.2005. http://www.vrco.com/CAVE_USER/.
- [26] Lawrence Chen, Hai Tao, Thomas Huang, Tsomuto Miyasato, and Ryohei Nakatsu. Emotion Recognition from Audiovisual Information. In *IEEE Second Workshop on Multimedia Signal Processing*, pages 83–88, December 1998.
- [27] Philip Cohen, David McGee, Sharon Oviatt, Lizhong Wu, Joshua Clow, Robert King, Simon Julier, and Lawrence Rosenblum. Multimodal Interactions for 2D and 3D Environments. *IEEE Computer Graphics and Applications*, pages 10–13, July/August 1999 1999.
- [28] Carolina Cruz-Neira, Daniel J. Sandin, and Thomas A. DeFanti. Surround-Screen Projection-Based Virtual Reality: The Design and Implementation of the CAVE. In *Proc. of ACM SIGGRAPH 93*, pages 135–142. ACM, 1993.
- [29] Mihaly Csikszentmihalyi. *Beyond Boredom and Anxiety: Experiencing Flow in Work and Play*. San Fransisco: Jossey-Bass, 1975.

- [30] Mihaly Csikszentmihalyi. *Creativity: Flow and the Psychology of Discovery and Invention*. New York: HarperCollins, 1996.
- [31] Mihaly Csikszentmihalyi and Isabella Selega Csikszentmihalyi. *Flow: The Psychology of Optimal Experience*, chapter Introduction to Part IV. New York: Cambridge University Press, 1988.
- [32] Sofia Dahl and Anders Friberg. Expressiveness of Musician's Body Movements in Performances on Marimba. In *Proceedings of the 2003 International Gesture Workshop*, pages 479–486, 2003.
- [33] Char Davies. *Women, Art, and Technology*, chapter Landscape, Earth, Body, Being, Space, and Time in the Immersive Virtual Environments Osmose and Ephémère, pages 322–337. The MIT Press, 2003.
- [34] Char Davies and John Harrison. Osmose: Towards Broadening the Aesthetics of Virtual Reality. *SIGGRAPH Comput. Graph.*, 30(4):25–28, 1996.
- [35] Gerwin de Haan, Michal Koutek, and Frits H. Post. IntenSelect: Using Dynamic Object Rating for Assisting 3D Object Selection. In *Virtual Environments 2005*, pages 201–209, 2005.
- [36] Joachim Deisinger, Roland Blach, Gerold Wesche, Ralf Breining, and Andreas Simon. Towards Immersive Modeling - Challenges and Recommendations: A Workshop Analyzing the Needs of Designers. In *Eurographics Workshop on Virtual Environments 2000*, pages 145–156, 2000.
- [37] Frank Dellaert, Thomas Polzin, and Alex Waibel. Recognizing Emotion in Speech. In *Proceedings of the Fourth International Conference on Spoken Language*, volume 3, pages 1970–1973, Pittsburgh, PA, USA, October 1996.
- [38] Paul Dourish. *Where the Action Is : The Foundations of Embodied Interaction*. The MIT Press, 2004.
- [39] Athanasios Drosopoulos, Themis Balomenos, Spiros Ioannou, Kostas Karpouzis, and Stefanos Kollias. Emotionally-rich Man-machine Interaction Based on Gesture Analysis. In *Universal Access in HCI: Inclusive Design in the Information Society*, pages 1372–1376, Crete, Greece, June 2003.
- [40] Nathaniel Duca, Peter D. Kirchner, and James T. Klosowski. Stream Caching: Optimizing Data Flow within Commodity Visualization Clusters. In *Proceedings of the Workshop on Commodity-Based Visualization Clusters*, 2002.
- [41] Matthew Willard Eldridge. *Designing graphics architectures around scalability and communication*. PhD thesis, 2001.
- [42] Norbert Elias. *Time: An Essay*. Blackwell Publishers, 1992.
- [43] Ronald Fedkiw, Jos Stam, and Henrik Wann Jensen. Visual Simulation of Smoke. In *Proceedings of ACM SIGGRAPH 2001*, pages 15–22. ACM Press/ACM SIGGRAPH, 2001.
- [44] Michele Fiorentino, Antonio Uva, and Giuseppe Monno. The SenStylus: A Novel Rumble-Feedback Pen Device for CAD Application in Virtual Reality. In *WSCG'2005 Full Papers*, pages 131–138, 2005.
- [45] Mark Foskey, Miguel Otaduy, and Ming Lin. ArtNova: Touch-Enabled 3D Model Design. In *Proc. of IEEE Virtual Reality 2002*, pages 119–126. IEEE Computer Society, 2002.
- [46] Nick Foster and Ronald Fedkiw. Practical Animation of Liquids. In *Proceedings of ACM SIGGRAPH 2001*, pages 23–30. ACM Press/ACM SIGGRAPH, 2001.

- [47] Anders Friberg and Roberto Bresin. Automatic Musical Punctuation: A Rule System and a Neural Network Approach. In *Proceedings of KANSEI - The Technology of Emotion AIMI International Workshop*, pages 159–163. AIMI - Associazione di Informatica Musicale Italiana, 1997.
- [48] Joseph Gabbard, Deborah Hix, and Edward Swan. User-Centered Design and Evaluation of Virtual Environments. *IEEE Computer Graphics and Applications*, 19(6):51–59, 1999.
- [49] Guy Garnett, Mangesh Jonnalagadda, Ivan Elezovic, Timothy Johnson, and Kevin Small. Technological Advances for Conducting a Virtual Ensemble. In *Proceedings of the International Computer Music Conference*, pages 167–169, 2001.
- [50] Guy Garnett, Fernando Malvar-Ruiz, and Fred Stoltzfus. Virtual Conducting Practice Environment. In *Proceedings of the 1999 International Computer Music Conference*, pages 371–374, Beijing, 1999. International Computer Music Association.
- [51] Brian Goldiez, Glenn Martin, Jason Daly, Donald Washburn, and Todd Lazarus. Software Infrastructure for Multi-Modal Virtual Environments. In *ICMI '04: Proceedings of the 6th international conference on Multimodal interfaces*, pages 303–308, New York, NY, USA, 2004. ACM Press.
- [52] Tovi Grossman, Daniel Wigdor, and Ravin Balakrishnan. Multi-Finger Gestural Interaction with 3D Volumetric Displays. In *UIST '04: Proceedings of the 17th annual ACM symposium on User interface software and technology*, pages 61–70, New York, NY, USA, 2004. ACM Press.
- [53] Stephen Haflich and Markus Burns. Following a Conductor: The Engineering of an Input Device. In *Proceedings of the International Computer Music Conference*, 1983.
- [54] Perttu Hämäläinen, Tommi Ilmonen, Johanna Höysniemi, Mikko Lindholm, and Ari Nykänen. Martial Arts in Artificial Reality. In *CHI '05: Proceeding of the SIGCHI conference on Human factors in computing systems*, pages 781–790, New York, NY, USA, 2005. ACM Press.
- [55] VTI Hamlin. VTI Hamlin Acceleration Sensors. Visited on 7.6.2006. URL: http://www.vti.fi/productsen/productsen_2.1-1.3.html.
- [56] Jennifer Healey and Rosalind Picard. Digital Processing of Affective Signals. In *Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing*, volume 6, pages 3749–3752, May 1998.
- [57] Jarmo Hiipakka, Rami Hänninen, Tommi Ilmonen, Hannu Napari, Tapio Lokki, Lauri Savioja, Hannu Huopaniemi, Matti Karjalainen, Tero Tolonen, Vesa Välimäki, Seppo Välimäki, and Tapio Takala. Virtual Orchestra Performance. In *Visual Proceedings of SIGGRAPH'97*, page 81, Los Angeles, 1997. ACM SIGGRAPH.
- [58] Jarmo Hiipakka, Tommi Ilmonen, Tapio Lokki, Matti Gröhn, and Lauri Savioja. Implementation Issues of 3D Audio in a Virtual Room. In *Proc. 13th Symposium of IS&T/SPIE, Electronic Imaging 2001*, volume 4297B (The Engineering Reality of Virtual Reality), pages 486–495, 2001.
- [59] Jarmo Hiipakka, Tommi Ilmonen, Tapio Lokki, and Lauri Savioja. Sound Signal Processing for a Virtual Room. In *Proceedings of the 10th Signal Processing Conference (EUSIPCO2000)*, pages 2221–2225, Tampere, Finland, 2000.
- [60] Michelle Hlubinka, Jennifer Beaudin, Emmanuel Munguia Tapia, and John S. An. AltarNation: Interface Design for Meditative Communities. In *CHI '02:*

- CHI '02 extended abstracts on Human factors in computing systems*, pages 612–613, New York, NY, USA, 2002. ACM Press.
- [61] Caroline Hummels and Pieter Jan Stappers. Meaningful Gestures for Human Computer Interaction: Beyond Hand Postures. In *Proceedings of the Third International Conference on Automatic Face and Gesture Recognition*, pages 591–596, April 1998.
- [62] Greg Humphreys, Matthew Eldridge, Ian Buck, Gordan Stoll, Matthew Everett, and Pat Hanrahan. WireGL: A Scalable Graphics System for Clusters. In *SIGGRAPH '01: Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, pages 129–140, New York, NY, USA, 2001. ACM Press.
- [63] Greg Humphreys, Mike Houston, Ren Ng, Randall Frank, Sean Ahern, Peter D. Kirchner, and James T. Klosowski. Chromium: A Stream-Processing Framework for Interactive Rendering on Clusters. In *SIGGRAPH '02: Proceedings of the 29th annual conference on Computer graphics and interactive techniques*, pages 693–702, New York, NY, USA, 2002. ACM Press.
- [64] Tommi Ilmonen. Immersive 3D User Interface for Computer Animation Control. In *The Proceedings of the International Conference on Computer Vision and Graphics 2002*, pages 352–359, Zakopane, Poland, September 2002.
- [65] Tommi Ilmonen. Immersive 3D User Interface for Particle System Control - AnimaLand. *Journal of Applied Mathematics and Computer Science*, 2003. To be published.
- [66] Tommi Ilmonen, Tapio Takala, and Juha Laitinen. Collision Avoidance and Surface Flow for Particle Systems Using Distance/Normal Grid. In Joaquim Jorge and Vaclav Skala, editors, *WSCG 2006 Full Papers Proceedings*, pages 79–86, 2006.
- [67] Tommi Ilmonen, Tapio Takala, and Juha Laitinen. Soft Edges and Burning Things: Enhanced Real-Time Rendering of Particle Systems. In Vaclav Skala, editor, *WSCG 2006 Short Papers Proceedings*, pages 33–38, 2006.
- [68] Hiroshi Ishii and Brygg Ullmer. Tangible Bits: Towards Seamless Interfaces between People, Bits and Atoms. In *CHI '97: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 234–241, New York, NY, USA, 1997. ACM Press.
- [69] Janne Jalkanen and Hannu Napari. How to Build a Virtual Room. In *Proc. 13th Symposium of IS&T/SPIE, Electronic Imaging 2001*, volume 4297B (The Engineering Reality of Virtual Reality), pages 475–485, Finland, 2001.
- [70] Blanchette Jasmin and Mark Summerfield. *C++ GUI Programming with Qt 3*. Prentice Hall, 2004.
- [71] Christopher Just, Allen Bierbaum, Albert Baker, and Carolina Cruz-Neira. VR Juggler: A Framework for Virtual Reality Development. In *Proceedings of the 2nd Immersive Projection Technology Workshop (IPT98)*, Iowa, May 1998.
- [72] R. S. Kalawsky, S. T. Bee, and S. P. Nee. Human Factors Evaluation Techniques to Aid Understanding of Virtual Interfaces. *BT Technology Journal*, 17(1):128–141, 1999.
- [73] Kostas Karpouzis, Amaryllis Raouzaoui, Athanasios Drosopoulos, Spiros Ioannou Themis Balomenos, Nicolas Tsapatsoulis, and Stefanos Kollias. Facial Expression and Gesture Analysis for Emotionally-rich Man-machine Interaction. In *3D Modeling and Animation: Synthesis and Analysis Techniques*, 2003.

- [74] Kulwinder Kaur. *Designing Virtual Environments for Usability*. PhD thesis, City University, London, 1998.
- [75] Fumio Kawakami, Motohiro Okura, Hiroshi Yamada, Hiroshi Harashima, and Shigeo Morishima. An Evaluation of 3-D Emotion Space. In *Proceedings of the 4th IEEE International Workshop on Robot and Human Comm. RO-MAN '95*, pages 269–274, Tokyo, Japan, July 1995.
- [76] Simeon Keates and Peter Robinson. The Use of Gestures in Multimodal Input. In *Assets '98: Proceedings of the third international ACM conference on Assistive technologies*, pages 35–42, New York, NY, USA, 1998. ACM Press.
- [77] Daniel Keefe, Daniel Feliz, Tomer Moscovich, David Laidlaw, and Joseph LaViola. CavePainting: A Fully Immersive 3D Artistic Medium and Interactive Experience. In *Proceedings on 2001 Symposium on Interactive 3D graphics*, pages 85–93. ACM Press New York, NY, USA, 2001.
- [78] Mark J. Kilgard. *OpenGL Programming for the X Window System*. Addison-Wesley Professional, 1996.
- [79] Eric Klein, Greg S. Schmidt, Erik B. Tomlin, and Dennis G. Brown. Dirt Cheap 3-D Spatial Audio. *Linux Journal*, pages 78–87, October 2005.
- [80] Christian Knöpfle. No Silver Bullet but Basic Rules – User Interface Design for Virtual Environments. In *Human-Computer Interaction: Theory and Practice (Part I)*, volume 1, pages 1158–1162, 2003.
- [81] Paul Kolesnik and Matcelo Wanderley. Recognition, Analysis and Performance with Expressive Conducting Gestures. In *Proceedings of the International Computer Music Conference*, 2004.
- [82] Myron W. Krueger. *Artificial Reality 2*. Addison-Wesley Professional, second edition, 1991.
- [83] Mikko Laakso. Comparison of Hand- and Wand Related Navigation in Virtual Environments. In *Human-Computer Interaction: Theory and Practice (Part I)*, volume 1, pages 1163–1167, 2003.
- [84] George Lakoff. *Women, Fire, and Dangerous Things*. University of Chicago Press, 1987.
- [85] Arnauld Lamorlette and Nick Foster. Structural Modeling of Flames for a Production Environment. In *Proceedings of the 29th annual conference on Computer graphics and interactive techniques*, pages 729–735. ACM Press, 2002.
- [86] James Landay and Brad Myers. Sketching Interfaces: Toward More Human Interface Design. *Computer*, 34(3):56–64, March 2001.
- [87] Jeff Lander. The Ocean Spray in Your Face. *Game Developer*, pages 13–19, July 1998.
- [88] Marc Erich Latoschik. A Gesture Processing Framework for Multimodal Interaction in Virtual Reality. In *AFRIGRAPH '01: Proceedings of the 1st international conference on Computer graphics, virtual reality and visualisation*, pages 95–100, New York, NY, USA, 2001. ACM Press.
- [89] Nigay Laurence and Coutaz Joëlle. A Design Space for Multimodal Systems: Concurrent Processing and Data Fusion. In *The proceedings of InterCHI '93, joint conference of ACM SIG-CHI and INTERACT*, pages 172–178, 1993.
- [90] Kevin Lefebvre. Exploration of the Architecture Behind HP's New Immersive Visualization Solutions, 2004.

- [91] Frederick Li, Rynson Lau, and Frederick Ng. Collaborative distributed virtual sculpting. In *Proc. of IEEE Virtual Reality 2001*, pages 217–224. IEEE Computer Society, 2001.
- [92] Tapio Lokki, Jarmo Hiipakka, Rami Hänninen, Tommi Ilmonen, Lauri Savioja, and Tapio Takala. Real-Time Audiovisual Rendering and Contemporary Audiovisual Art. *Organised Sound*, 3(3):219–233, 1998.
- [93] Tapio Lokki, Tommi Ilmonen, Wille Mäkelä, and Tapio Takala. Upponurkka: An Inexpensive Immersive Display for Public Vr Installations. In *EEE Virtual Reality 2006, Workshop on Emerging Display Technologies*, pages 15–18, 2006.
- [94] Tapio Lokki, Lauri Savioja, Jari Huopaniemi, Rami Hänninen, Tommi Ilmonen, Jarmo Hiipakka, Ville Pulkki, Riitta Väänänen, and Tapio Takala. Virtual Concerts in Virtual Spaces - in Real Time (invited paper). In *the CDROM of the ASA/EAA/DEGA Joint Meeting*, Berlin, Germany, Mar. 14-19 1999. paper n. 1pSP1. Abstract in *JASA* 105(2), pp. 979 and *ACUSTICA* 85(Suppl.1),pp. S53.
- [95] Wille Mäkelä and Tommi Ilmonen. Drawing, Painting and Sculpting in the Air. Development Studies about an Immersive Free-Hand Interface for Artists. In *IEEE VR 2004 Workshop Proceedings*, pages 89–92, 2004.
- [96] Wille Mäkelä, Markku Reunanen, Tapio Takala, and Tommi Ilmonen. Possibilities and Limitations of Immersive Free-Hand Expression: a Case Study with Professional Artists. In *ACM Multimedia 2004 Conference*, pages 504–507, October 2004.
- [97] Alberto Manguel. *The History of Reading*. Penguin, 1997.
- [98] Jennifer Mankoff, Scott E. Hudson, and Gregory D. Abowd. Providing Integrated Toolkit-Level Support for Ambiguity in Recognition-Based Interfaces. In *Proceedings of the CHI 2000 conference on Human factors in computing systems*, pages 368–375, The Hague, The Netherlands, 2000. ACM Press New York, NY, USA.
- [99] Teresa Marrin. Toward an Understanding of Musical Gesture: Mapping Expressive Intention with the Digital Baton. Master’s thesis, MIT, 1996.
- [100] Teresa Marrin and Rosalind Picard. The ”Conductors Jacket”: A Device for Recording Expressive Musical Gestures. In *Proceedings of the International Computer Music Conference*, pages 215–219, Ann Arbor, Michigan USA, October 1998. International Computer Music Association.
- [101] Dominic W. Massaro. A Framework for Evaluating Multimodal Integration by Humans and a Role for Embodied Conversational Agents. In *ICMI ’04: Proceedings of the 6th international conference on Multimodal interfaces*, pages 24–31, New York, NY, USA, 2004. ACM Press.
- [102] M. Mathews. The Radio Baton and Conductor Program, or: Pitch, the Most Important and Least Expressive Part of Music. *Computer Music Journal*, 19(4), 1991.
- [103] Max Mathews. The Conductor Program. In *Proceedings of the International Computer Music Conference*, 1976.
- [104] David McAllister. The Design of an API for Particle Systems. Technical report, University of North Carolina, January 2000.
- [105] Brock McElheran. *Conducting Technique for Beginners and Professionals*. Oxford University Press, Oxford/New York, 1989.

- [106] Mark R. Mine, Jr. Frederick P. Brooks, and Carlo H. Sequin. Moving Objects in Space: Exploiting Proprioception in Virtual-Environment Interaction. In *SIGGRAPH '97: Proceedings of the 24th annual conference on Computer graphics and interactive techniques*, pages 19–26, New York, NY, USA, 1997. ACM Press/Addison-Wesley Publishing Co.
- [107] Brian C.J. Moore. *An Introduction to the Psychology of Hearing*. Academic Press, fifth edition, 2003.
- [108] Hideyuki Morita, Shuji Hashimoto, and Sadamu Ohteru. Computer Music System that Follows a Human Conductor. In *Proceedings of the 1989 International Computer Music Conference*, pages 207–210, San Francisco, 1989. International Computer Music Association.
- [109] Hideyuki Morita, Hiroshi Watanabe, Tsutomu Harada, Sadamu Ohteru, and Shuji Hashimoto. Knowledge Information Processing in Conducting Music Performer. In *Proceedings of the International Computer Music Conference*, San Francisco, 1990. Computer Music Association.
- [110] Declan Murphy, Tue Haste Andersen, and Kristoffer Jensen. Conducting Audio Files via Computer Vision. In *Proceedings of the 2003 International Gesture Workshop*, pages 529–540, 2003.
- [111] Martin Naef, Edouard Lamoray, Oliver Staadt, and Markus Gross. The Blue-C Distributed Scene Graph. In *EGVE '03: Proceedings of the workshop on Virtual environments 2003*, pages 125–133, New York, NY, USA, 2003. ACM Press.
- [112] Duc Quang Nguyen, Ronald Fedkiw, and Henrik Wann Jensen. Physically Based Modeling and Animation of Fire. In *Proceedings of the 29th annual conference on Computer graphics and interactive techniques*, pages 721–728. ACM Press, 2002.
- [113] Michael Nielsen, Moritz Störring, Thomas Moeslund, and Erik Granum. A Procedure for Developing Intuitive and Ergonomic Gesture Interfaces for HCI. In *Proceedings of the 2003 International Gesture Workshop*, pages 409–420, 2003.
- [114] Donald A. Norman. *The Invisible Computer: Why Good Products Can Fail, the Personal Computer Is So Complex, and Information Appliances Are the Solution*. The MIT Press, 1998.
- [115] Eric Olson. Cluster Juggler - PC Cluster Virtual Reality. Master's thesis, Iowa State University, 2002.
- [116] Sharon Oviatt. TeN Myths of Multimodal Interaction. *Commun. ACM*, 42(11):74–81, 1999.
- [117] Stephen E. Palmer. *Vision Science: Photons to Phenomenology*. The MIT Press, 1999.
- [118] Rosalind Picard. *Affective Computing*. The MIT Press, Cambridge, Massachusetts, 1997.
- [119] Rosalind W. Picard. Affective Computing. Technical Report 321, MIT Media Laboratory Perceptual Computing Section, 1995.
- [120] Márcio S. Pinho, Doug A. Bowman, and Carla M.D.S. Freitas. Cooperative Object Manipulation in Immersive Virtual Environments: Framework and Techniques. In *VRST '02: Proceedings of the ACM symposium on Virtual reality software and technology*, pages 171–178, New York, NY, USA, 2002. ACM Press.

- [121] Miller Puckette. Pure Data: Another Integrated Computer Music Environment. In *Proceedings of the Second Intercollege Computer Music Concerts*, pages 37–41, 1996.
- [122] Miller Puckette. Max at Seventeen. *Computer Music Journal*, (4):31–43, 2002.
- [123] Ville Pulkki. Virtual Sound Source Positioning Using Vector Base Amplitude Panning. *Journal of the Audio Engineering Society*, 45(6):456–466, June 1997.
- [124] William T. Reeves. Particle Systems — a Technique for Modeling a Class of Fuzzy Objects. In *Proceedings of the 10th annual conference on Computer graphics and interactive techniques*, pages 359–375, 1983.
- [125] William T. Reeves and Ricki Blau. Approximate and Probabilistic Algorithms for Shading and Rendering Structured Particle Systems. In *Proceedings of the 12th annual conference on Computer graphics and interactive techniques*, pages 313–322. ACM Press, 1985.
- [126] Gerhard Reitmayr and Dieter Schmalstieg. An Open Software Architecture for Virtual Reality Interaction. In *Proceedings of the ACM symposium on Virtual reality software and technology*, pages 47–54. ACM Press New York, NY, USA, 2001.
- [127] Markku Reunanen, Karri Palovuori, Tommi Ilmonen, and Wille Mäkelä. Näprä — Affordable Fingertip Tracking with Ultrasound. In *Virtual Environments 2005*, pages 51–58, 2005.
- [128] Craig W. Reynolds. FlockS, Herds and Schools: A Distributed Behavioral Model. In *Proceedings of the 14th annual conference on Computer graphics and interactive techniques*, pages 25–34. ACM Press, 1987.
- [129] Deb Roy and Alex Pentland. AutomAtic Spoken Affect Classification and Analysis. In *Proceedings of the Second International Conference on Automatic Face and Gesture Recognition*, pages 363–367, October 1996.
- [130] Max Rudolf. *The Grammar of Conducting: A Comprehensive Guide to Baton Technique and Interpretation*. Wadsworth Publishing, third edition, 1995.
- [131] Daniel Salber, Anind K. Dey, and Gregory D. Abowd. The Context Toolkit: Aiding the Development of Context-Enabled Applications. In *Proceeding of the CHI 99 Conference on Human factors in Computing Systems*, pages 434–441, Pittsburgh, Pennsylvania, United States, 1999. ACM Press New York, NY, USA.
- [132] Benjamin Schaeffer and Camille Goudeseune. Syzygy: Native PC Cluster VR. In *Proceedings of IEEE Virtual Reality 2003*, pages 15–22, 2003.
- [133] Robert Schalkoff. *Pattern Recognition: Statistical, Structural, and Neural Approaches*. J. Wiley, 1992.
- [134] Jocelyn Scheirer, Raul Fernandez, and Rosalind Picard. Expression Glasses: A Wearable Device for Facial Expression Recognition. In *Proceedings of CHI*. ACM SIGCHI, may 1999.
- [135] Steven Schkolne, Michael Pruett, and Peter Schröder. SurfAce Drawing: Creating Organic 3D Shapes With the Hand and Tangible Tools. In *CHI '01: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 261–268. ACM Press, 2001.
- [136] L. Schoemaker, J. Nijtmans, A. Camurri, F. Lavagetto, P. Morasso, C. Benoit, T. Guiard-Marigny, B. Le Goff, J. Robert-Ribes, A. Adjoudani, I. Deféé,

- S. Münch, K. Hartung, and J. Blauert. A Taxonomy of Multimodal Interaction in the Human Information Processing System. Technical report, ESPRIT BRA, No. 8579, 1995.
- [137] Jakub Segen, Aditi Majumder, and Joshua Gluckman. Virtual Dance and Music Conducted By a Human Conductor. *Eurographics*, 19(3), 2000.
- [138] Karl Sims. Particle Animation and Rendering Using Data Parallel Computation. In *SIGGRAPH '90: Proceedings of the 17th annual conference on Computer graphics and interactive techniques*, pages 405–413, New York, NY, USA, 1990. ACM Press.
- [139] Michael Siracusa, Louis-Philippe Morency, Kevin Wilson, John Fisher, and Trevor Darrell. A Multi-Modal Approach for Determining Speaker Location and Focus. In *ICMI '03: Proceedings of the 5th international conference on Multimodal interfaces*, pages 77–80, New York, NY, USA, 2003. ACM Press.
- [140] Oliver G. Staadt, Justin Walker, Christof Nuber, and Bernd Hamann. A Survey and Performance Analysis of Software Platforms for Interactive Cluster-Based Multi-Screen Rendering. In *EGVE '03: Proceedings of the workshop on Virtual environments 2003*, pages 261–270, New York, NY, USA, 2003. ACM Press.
- [141] Jos Stam. Stable Fluids. In *Proceedings of ACM SIGGRAPH 99*, pages 121–128. ACM Press/ACM SIGGRAPH, 1999.
- [142] Jos Stam. Interacting with Smoke and Fire in Real Time. *Commun. ACM*, 43(7):76–83, 2000.
- [143] Jos Stam. Flows on Surfaces of Arbitrary Topology. *ACM Trans. Graph.*, 22(3):724–731, 2003.
- [144] Jos Stam and Eugene Fiume. Depicting Fire and Other Gaseous Phenomena Using Diffusion Processes. In *SIGGRAPH '95: Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*, pages 129–136, New York, NY, USA, 1995. ACM Press.
- [145] Gordon Stoll, Matthew Eldridge, Dan Patterson, Art Webb, Steven Berman, Richard Levy, Chris Caywood, Milton Taveira, Stephen Hunt, and Pat Hanrahan. Lightning-2: A High-Performance Display Subsystem for PC Clusters. In *SIGGRAPH '01: Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, pages 141–148, New York, NY, USA, 2001. ACM Press.
- [146] Alexander Streit, Ruth Christie, and Andy Boud. Understanding Next-Generation VR: Classifying Commodity Clusters for Immersive Virtual Reality. In *GRAPHITE '04: Proceedings of the 2nd international conference on Computer graphics and interactive techniques in Australasia and South East Asia*, pages 222–229, New York, NY, USA, 2004. ACM Press.
- [147] Tapio Takala, Rami Hänninen, Vesa Välimäki, Lauri Savioja, Jyri Huopaniemi, Tommi Huutilainen, and Matti Karjalainen. An Integrated system for Virtual Audio Reality. In *Preprint 4229 (M-4), Copenhagen: 100th AES Convention of Acoustic Engineering Society: 46*, Copenhagen, Denmark, 1996.
- [148] Russel M. Taylor, Thomas C. Hudson, Adam Seeger, Hans Weber, Jeffrey Juliano, and Aron T. Helser. VRPN: A Device-Independent, Network-Transparent VR Peripheral System. In *Proceedings of the ACM symposium on Virtual reality software and technology*, pages 55–61. ACM Press New York, NY, USA, 2001.

- [149] Forrest Tobey. The Ensemble Member and the Conducted Computer. In *Proceedings of the 1995 International Computer Music Conference*, pages 529–530, San Francisco, 1995. Computer Music Association.
- [150] Forrest Tobey and Ichiro Fujinaga. Extraction of Conducting Gestures in 3D Space. In *Proceedings of the International Computer Music Conference*, pages 305–307, San Francisco, 1996. Computer Music Association.
- [151] Gerard J. Tortora and Sandra R. Grabowski. *Introduction to the Human Body: The Essentials of Anatomy and Physiology*. Wiley, 2004.
- [152] S. Usa and Y. Mochida. A Multi-Modal Conducting Simulator. In *Proceedings of the International Computer Music Conference*, pages 25–32, Ann Arbor, Michigan USA, October 1998. International Computer Music Association.
- [153] Martin Usoh, Kevin Arthur, Mary C. Whitton, Rui Bastos, Anthony Steed, Mel Slater, and Jr. Frederick P. Brooks. Walking > Walking-in-Place > Flying, in Virtual Environments. In *SIGGRAPH '99: Proceedings of the 26th annual conference on Computer graphics and interactive techniques*, pages 359–364, New York, NY, USA, 1999. ACM Press/Addison-Wesley Publishing Co.
- [154] John van der Burg. Building an Advanced Particle System. *Game Developer*, March 2000.
- [155] Bradley Vines, Marcelo Wanderlay, Carol Krumhansl, Regina Nuzzo, and Daniel Levitin. Performance Gestures of Musicians: What Structural and Emotional Information Do They Convey ? In *Proceedings of the 2003 International Gesture Workshop*, pages 468–478, 2003.
- [156] Norman G. Vinson. Design Guidelines for Landmarks to Support Navigation in Virtual Environments. In *CHI '99: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 278–285, New York, NY, USA, 1999. ACM Press.
- [157] Stefan Waldherr, Roseli Romero, and Sebastian Thrun. A Gesture Based Interface for Human-Robot Interaction. *Auton. Robots*, 9(2):151–173, 2000.
- [158] John Walker, editor. *The Autodesk File - Bits of History, Words of Experience*. fourth edition, 1994. URL: <http://www.fourmilab.ch/autofile/>.
- [159] Brian A. Wandell. *Foundations of Vision*. Sinauer Associates, 1995.
- [160] Mark Weiser. The Computer for the 21st Century. *SIGMOBILE Mob. Comput. Commun. Rev.*, 3(3):3–11, 1999.
- [161] Jakub Wejchert and David Haumann. Animation Aerodynamics. *ACM SIGGRAPH Computer Graphics*, 25:19–22, July 1991.
- [162] Gerold Wesche and Marc Droske. Conceptual Free-Form Styling on The Responsive Workbench. In *VRST '00: Proceedings of the ACM symposium on Virtual reality software and technology*, pages 83–91, New York, NY, USA, 2000. ACM Press.
- [163] Carsten Winkelholz and T. Alexander. Approach for Software Development of Parallel Real-Time VE Systems on Heterogenous Clusters. In *EGPGV '02: Proceedings of the Fourth Eurographics Workshop on Parallel Graphics and Visualization*, pages 23–32, Aire-la-Ville, Switzerland, Switzerland, 2002. Eurographics Association.
- [164] Maththew Wright, Adrian Freed, and Ali Momeni. OpenSound Control: State of the Art 2003. In *Proceedings of the 2003 Conference on New Interfaces for Musical Expression (NIME-03)*, pages 153–159, Montreal, Canada, 2003.

- [165] Mike Wu and Ravin Balakrishnan. Multi-Finger and Whole Hand Gestural Interaction Techniques for Multi-User Tabletop Displays. In *UIST '03: Proceedings of the 16th annual ACM symposium on User interface software and technology*, pages 193–202, New York, NY, USA, 2003. ACM Press.
- [166] Toyotoshi Yamada, Hideki Hasimoto, and Naoko Tosa. Pattern Recognition of Emotion with Neural Network. In *Proceedings of the International Conference on Industrial Electronics, Control, and Instrumentation*, volume 1, pages 183–187, November 1995.
- [167] Jian Yang, Jiaoying Shi, Zhefan Jin, and Hui Zhang. Design and Implementation of a Large-Scale Hybrid Distributed Graphics System. In *EGPGV '02: Proceedings of the Fourth Eurographics Workshop on Parallel Graphics and Visualization*, pages 39–49, Aire-la-Ville, Switzerland, 2002. Eurographics Association.
- [168] Gary Yngve, James O'Brien, and Jessica Hodgins. Animating Explosions. In *Proceedings of ACM SIGGRAPH 2000*, pages 29–36. ACM Press/ACM SIGGRAPH, 2000.

HELSINKI UNIVERSITY OF TECHNOLOGY
PUBLICATIONS IN TELECOMMUNICATIONS SOFTWARE AND MULTIMEDIA

- TML-A1 Håkan Mitts
Architectures for wireless ATM
- TML-A2 Pekka Nikander
Authorization in agent systems: Theory and practice
- TML-A3 Lauri Savioja
Modeling techniques for virtual acoustics
- TML-A4 Teemupekka Virtanen
Four views on security
- TML-A5 Tapio Lokki
Physically-based auralization — Design, implementation, and evaluation
- TML-A6 Kari Pihkala print
Extensions to the SMIL multimedia language
- TML-A7 Kari Pihkala pdf
Extensions to the SMIL multimedia language
- TML-A8 Harri Kiljander
Evolution and usability of mobile phone interaction styles
- TML-A9 Leena Eronen
User centered design of new and novel products: case digital television
- TML-A10 Sanna Liimatainen and Teemupekka Virtanen (eds.)
NORDSEC 2004, Proceedings of the Ninth Nordic Workshop on Secure IT Systems
- TML-A11 Timo Aila
Efficient algorithms for occlusion culling and shadows
- TML-A12 Pablo Cesar
A graphics software architecture for high-end interactive TV terminals
- TML-A13 Samuli Laine
Efficient Physically-Based Shadow Algorithms
- TML-A14 Istvan Beszteri
Dynamic Layout Adaptation of Web Documents

printed version:

ISBN-13 978-951-22-8550-1

ISBN-10 951-22-8550-9

ISSN 1456-7911

electronic version:

ISBN-13 978-951-22-8551-8

ISBN-10 951-22-8551-7

ISSN 1455-9722