KUNGL TEKNISKA HÖGSKOLAN
Royal Institute of Technology
Numerical Analysis and Computing Science

CID-81, KTH, Stockholm, Sweden May 1998

Navigation and Devices

John Bowers, Monika Fleischmann, Sten-Olof Hellström, Michael Hoch,
Kai-Mikael Jää-Aro, Thomas Kulessa, Jasminko Novak, Jeffrey Shaw,
Wolfgang Strauss

CID
Centre for
User Oriented IT Design

Author: John Bowers, Monika Fleischmann, Sten-Olof Hellström, Michael Hoch, Kai-Mikael Jää-Aro, Thomas Kulessa, Jasminko Novak, Jeffrey Shaw, Wolfgang Strauss

Title: Navigation and Devices Report number: CID-81

ISSN number: ISSN 1403-0721

Publication date: May 1998

E-mail of author: bowers@nada.kth.se

URL of author: http://www.nada.kth.se/erena

Reports can be ordered from:

CID, Centre for User Oriented IT Design
Nada, Dept. Computing Science
KTH, Royal Institute of Technology
S-100 44 Stockhom, Sweden
telephone: + 46 8 790 91 00
fax: + 46 8 790 90 99
e-mail: cid@nada.kth.se
URL: http://www.nada.kth.se/cid/

eRENA

Deliverable 6.1

Navigation and Devices

ABSTRACT

Task 6.1 is concerned with developing new interfaces and new metaphors for more physical interaction with virtual environments, involving the entire body and its physical properties.

The deliverable is divided into three parts:

- “A Characterization of Input Devices used in Interactive Installations” develops a taxonomy of how input devices and space have been used in interactive installations.
- “Navigation for the Senses”describes several devices for whole-body interaction developed or under development at GMD.
- “Some Elementary Gestural Techniques for Real-Time Interaction in Artistic Performances” describes gesture-based interfaces for multimedia performances.

Document	eRENA-D6.1
Type	Deliverable report with video material on eRENA tape
Status	Final
Version	1.0
Date	30th May 1998
Author(s)	John Bowers (KTH), Monika Fleischmann (GMD), Sten-Olof Hellström (KTH), Michael Hoch (ZKM), Kai-Mikael Jää-Aro (KTH), Thomas Kulessa (GMD), Jasminko Novak (GMD), Jeffrey Shaw (ZKM), Wolfgang Strauss (GMD)
Task	6.1

DELIVERABLE 6.1
Navigation and Devices

Part II
Navigation for the Sea

Part III

Overview

Task 6.1 is concerned with developing new interfaces and new metaphors for more physical interaction with virtual environments, involving the entire body and its physical properties.

The deliverable is divided into three parts:

- “A Characterization of Input Devices used in Interactive Installations”, by Michael Hoch and Jeffrey Shaw at ZKM, develops a taxonomy of how input devices and space have been used in interactive installations.
- “Navigation for the Senses”, by Jasminko Novak, Monika Fleischmann, Wolfgang Strauss and Thomas Kulessa at GMD, describes several devices for whole-body interaction developed or under development at GMD.
- “Some Elementary Gestural Techniques for Real-Time Interaction in Artistic Performances”, by John Bowers and Sten-Olof Hellström at KTH, describes the gesture-based interfaces they have developed for multimedia performances.

Fullfilment of promises

A number of goals have been set up for Task 6.1 (eRENA Project Programme, p 58).

- *The development of virtual environments for full-body interaction.* The MARS External User Interface Driver and Simple Gesture Interface Driver described in part II, section 4.4, are a software infrastructure for such virtual environments. The devices mentioned below have been developed with the express purpose of full-body interaction.
- *A multi-user interface for the Virtual Balance.* The EUID allows an arbitrary number of interface devices to be connected to a VRML browser, and this will be used to construct a forthcoming multi-VB application, as described in part II, section 3.3.
- *The theremin will be adapted to use as a computer interface.* Several simple theremin (and theremin-like) devices have been constructed and are being tested for use in various installations, as described in part II, section 4, and part III.
- *A balance sensitive floor for the CAVE.* Some of the theremins mentioned above have been used to create a floor which detects the presence of people and functions of their body posture, as described in part II, section 4.4.
- *A characterisation of interaction devices* Part I is an overview of a large number of interaction devices used in artistic installations and how these not only support interaction with the art pieces, but actively shape the viewer's experience of these.

In addition gestural interaction is considered in part I, section 5, part II, section 4.3 and part III and camera tracking in part I, section 5 and part II, section 5.

Ties to other work packages

Task 6.1 has ties to several other tasks. All tasks within WP 1 and task 3.1 contain aspects of interaction with installations, which can be informed by part I of this document. The extended galleries of Task 1.1 use the software platform (MARS EUID) described in part II, section 4.4. The multi-media performance in task 2.3 is also concerned with fluid gestural interaction, and the resulting interfaces are described in part III.

Part I

Survey of input and tracking devices used in artists' interactive installations at the ZKM.

Michael Hoch, Jeffrey Shaw

1. Introduction

In general an interface can be seen as an entity that forms a common boundary between two things. In terms of software it is a program that allows the user to interact with the system, in terms of hardware it is the associated circuitry that links one device with another. Due to [3] interaction devices can be categorized by means of locator, pick, keyboard, valuator and choice devices. The usefulness of a particular device depends on the interaction task that need to be performed and the interaction technique used. The most common interaction techniques that have been used particularly with graphical interactive systems and that are proven to be useful are direct manipulation, iconic user interfaces and the WYSIWYG principle (what you see is what you get). Most of these paradigms have been used with traditional desktop and menu based systems. Within these system mostly a combination of different techniques is more appropriate than a consistent use of one single paradigm. As an addition, the use of space can facilitate the use of the computer due to the human spatial memory skills. Unfortunately, the use of space is mostly limited to the space on the monitor or a projection screen. This space is in fact limited to the size of the screen, so that often new Information will substitute the old. The user has to memorize entities by means of context alone, the spatial relationship to the place in space is lost.

In this paper, we will try to explore different categories in a somewhat different approach. Artists use of interfaces often show a variety and quality not found in industrial applications. We will therefore explore some of the interactive installations at the ZKM in Karlsruhe. Over the last decade artists who have been making computer based interactive works have (often unconsciously) identified many basic paradigms of person machine interaction. We will describe the idiosyncratic and uniqueness of artists use of input device that often lead to innovative or interesting strategies. We will first explore how artists make use of generic devices to effectively transform them for their needs and give it some added value. Thereafter, we explore some metaphorical or symbolic devices, vehicular devices, and, finally, some gestural devices. This overview of artists practice can give hints as a departure for future development and use of input and interaction devices.

2. Generic Devices

In this section we describe the devices that have been used by artists in a generic way, i.e. devices that have been used in a rather traditional or common way. We try to point out how the artists succeed in transforming the interface in a way that is suitable for the application or in a way that it is no longer perceived as such.

2.1 Fruit Machine (Agnes Hegedüs, 1991)

Device:

Three 3D-Joysticks

Setup:

Three metal poles with 3D-Joysticks are located in front of a projection screen. The projection shows three parts of an octagonal form. Coordination of all three users is necessary to fit the parts to a single form, which then results in a virtual money output on the screen.

Users:

Three users interact simultaneously to get a meaningful interaction. If there are less then three users, it is left to the users to figure out that three are needed.

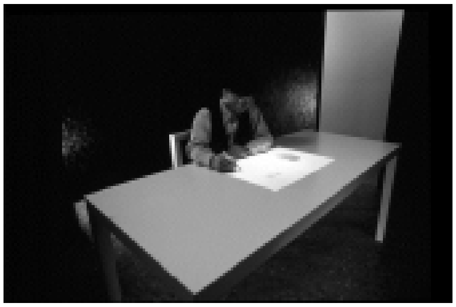
Transform:

The main theme of this installation is the need for cooperation between three users to get a desired output. The users have to coordinate their interactions to fit the three parts to a single form, which shows the difficulty of three people working together. The generic device used here, a 3D-Joystick to control the single parts, is used as a controller for the puzzle. The Joystick itself is not easy to use and it takes some time for the novice user to figure out that cooperation among the others is needed. This reflects the theme of the installation and therefore reinforces the storyline.

Use of Space:

The use of space is limited to the arrangement of the three poles in front of the projection screen: they are set up in a row so that the user is aware of others participating in the experience. By placing the poles rather close next to each other, human communication is possible during the interaction and after the users succeed. Nevertheless, the main focus of attention is drawn to the projection screen and the displayed form.

2.2 Beyond Pages (Masaki Fujihata, 1995)



Device:

Wacom Digitizer Tablet A2 integrated in table, wireless pen

Setup:

The user enters a room with chair and table. He interacts with a digital book that is projected on a Wacom Tablet. The tablet is integrated in the table and, hence, not visible to the user. The pen is used for turning pages, interacting with pages contents, and for triggering events in real space (switching of a lamp, starting a video sequence at a door).

Users:

There is one main user interacting with the book and sitting down at the table. But, the audience takes part in the experience because of the special environmental setup of the installation.

Transform:

By integrating the tablet into the table and projecting content on it, the generic tablet is not visible for the user. The pen is used in a generic way for triggering events, but the whole setup and environment that is created will let the user interact with the digital book and the objects themselves in a direct and intuitive way. Here, the special setup both in hardware and software lets the user perceive a pleasing environment, the generic device is seamlessly integrated in the environment and not such. The pen as a generic device gets a different meaning (as opposed to a traditional use calling up menus and pushing buttons), it gains some additional power while the user is interacting with the objects and the environment.

Use of Space:

Space is used by means of the room setup that integrates the projection table and real live objects like lamp and door. Furthermore, by using a projection on the table the audience can participate in the interactive experience because of the spatial awareness of objects in the environment, i.e. objects, user and potential visitors (audience) are situated in the same environment. Therefore, the user will be immersed as soon as he enters the room and, hence, the distance between the user and the main interaction device (the table) is made small, i.e. the user is invited to participate.

2.3 Liquid Views (M. Fleischmann, W. Strauss, CA. Bohn, 1993)

Device:

Touchscreen

Setup:

A pedestal with an embedded monitor stands in front of a large screen. The observer bend over the horizontal video picture and releases an alteration of the original picture through his own movements and through touching the surface. A spring conceived as a well, filled with virtual water, reflects our image. The world behind the mirror is regarded as untouchable; here, computer technology make it possible to create an interface which enables one to communicate with a virtual world of reflections.

Users:

Single user interface.

Transform:

By integrating the touchscreen in the pedestal and by displaying virtual water, the touchscreen is becoming invisible to the user. Touching the screen transforms to touching water or touching the mirrored image.

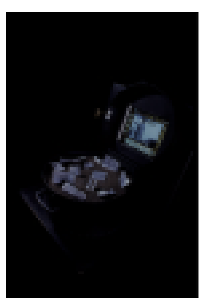
Use of Space:

The special setup that allows the user to bend over the horizontal video picture lets the user experience a real life situation. The way the user is situated in this environment makes up much of the success and intuitiveness of the installation.

3. Metaphoric & Symbolic Devices

In this section we will explore some examples of interface use that have a strong symbolic meaning in the way they are integrated in the environment or in the way they are used. The metaphorical meaning of such interfaces creates a specific feeling on the users site that is essential for the interactive experience and the quality of the art work.

3.1 Frontiers of Utopia (Jill Scott, 1995)

Device:

4 touch screens, suitcases with sculptural miniatures in metal, which are touchable with a key.

Setup:

In the corners of a dark, closed room are four units consisting of a monitor with a touch screen, a sculptural interface in the shape of a suitcase, and a projection surface. The suitcase in front of the monitor screen contains objects made of metal. When these objects are touched with a key the scenes change.

Users:

4 to five users can interact at simultaneously, though, each user has his own field of view, because the monitors and the interfaces are separated in 4 corners.

Transform:

In this work the spatial threshold between real and virtual scenery is emphasized. The suitcases are material reminders of this split, to which the time journey through dialog can be attached. The observer of her installation is not simply left alone in the conditions of a virtual world, but is able to test the tension between the virtual space of the story and the real space of daily experience. Touching the sculptural miniatures by using the key has a strong metaphoric impact: The user is not just triggering events like clicking buttons on the screen with a mouse, but she is captured through the symbolic expression of the setup with suitcases and metal figures.

Use of Space:

Here, real space is mainly used in conjunction with real, daily life objects to create a tension and a symbolic meaning of the interface.

3.2 Surprising Spiral (Ken Feingold, 1991)

Device:

fake book with touch screen, plastic mouth with sensor

Setup:

A rostrum is in front of a projection screen. Some steps lead to a table, upon which a big, book-shaped box, some fake books, and an oversized plastic mouth are fixed. These objects form a sculptural interface. The observer sits on a bench in front of the table and selects film and audio sequences that can partially be controlled by the user. When touching the mouth voices will be heard. They stop when releasing the mouth. The book has a glass plate (touch screen) with finger prints that serve as buttons. Pressing theses buttons will eventually alter the video sequences.

Users:

Single user interface.

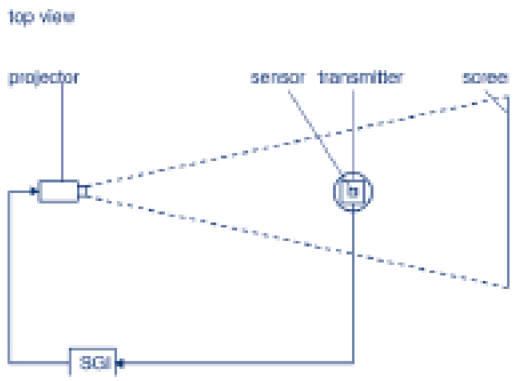
Transform:

The setup and the chosen devices (book with finger prints, mouth) in this piece has mostly a metaphoric or symbolic meaning. Like the mouth as a symbol for talking or the finger prints that are buttons itself. The artist deals with these symbols in a specific way giving the user only limited control over the system. An internal logic determines which sequences are actually to play. The input of the user is only part of this selection process leaving the user on the outside. The metaphoric relationship to the piece is the only relationship the user can have and the user is left with this experience.

Use of Space:

Space is used to emphasize the symbolic meaning, i.e. by using the sculptural interface and a bench setup for the user as well as placing the interface two steps up the ground on a platform.

3.3 Handsight (Agnes Hegedüs, 1991-93)



Device:

Polhemus 3D-tracker in eye shaped ball, plexi sphere on a plinth,

Setup:

A hand-held “eyeball” interface with an Polhemus sensor tracks the users hand position within a transparent sphere with an iris-like opening for the hand. These elements are accompanied by a round projection of an eye. Once the user penetrates inside the empty transparent globe the projected eye on screen opens into a virtual world. Using the hand-held eye, the user can navigate in this world.

Users:

Single user interface.

Transform:

The three elements in this installation are metaphors for the eye. The eye stands a surface where both, exterior reality and interior subjectivity can be reflected. By shaping the interface as an eye itself, this relationship, on one hand, becomes obvious for the user and, on the other hand, it creates a strong but subtle tension while the user is holding an eye in her hand to control the virtual camera. Thus, the eye becomes her own eye and a metaphor for perception. Using this strong metaphoric approach, the interface becomes intuitive to use, the functionality of the virtual camera, for example, need not to be explained.

Use of Space:

In this installation the spatial layout is given by the interaction devices. The user, holding the eye, finds himself within an eye (the plexi sphere), but is also present as an external observer.

3.4 Interactive Plant Growing (Christa Sommerer, Laurent Mignonneau, 1992)

Device:

5 plants with low voltage sensors

Setup:

In a room there is a large projection screen and plinths which have been distributed throughout. They have preserved plants on them which, when touched, send impulses to a computer via a sensory mechanism. Depending on its intensity, as a consequence of the touch, different types of growing plants are seeded and projected on the screen. The simulation of the growing plants is interrupted by touching a cactus.

Users:

Multiple user interface. Ideally there is one user at each plant, so that the single user gains “control” over the plant. But, more than one user can touch the plant simultaneously.

Transform:

The use of plants has a strong metaphoric meaning that is used here in a direct way, i.e. real plants are used to grow digital plants. The relationship is obvious. The mechanism itself, touching the plant to induce a seed in the virtual world, is itself not intuitive because it is unfamiliar. On the other hand, once the user knows about the controls it then becomes intuitive and creates a strong sensational feeling on the users site: Digital plants become touchable. The use of the cactus for triggering a *clear screen* operation, transforms the cactus to a role of device it naturally does not inheritates. But, touching a real cactus too hard will hurt your hand. The cactus here, is a metaphor for destruction in this way of meaning.

Use of Space:

The spatial setup allows multiple user to interact with a projection screen through the plants. Interaction between the different user is also possible because the user is aware of the interactions that take place left and right of “his” plant.

4. Vehicular Devices

In this section we will explore some devices that are used for navigation in virtual space. A special focus here will be so called vehicular devices. These devices do have the notion of movement build into the device itself. Some devices might be so familiar that the user would immediately understand the purpose of it, some other device might be placed in an environment in such a way that it can intuitively be used for navigational purposes.

4.1 Legible City (Jeffrey Shaw, 1988-91)

Device:

modified bicycle

Setup:

A bicycle with a small monitor on the handlebars is mounted in front of a big projection screen. When the observer pedals, a projection is activated and he can move through three different, simulated representations of cities (Manhattan, Amsterdam, and Karlsruhe). The architectural landscape of the streets is formed by letters and texts. Ground plans of the city can be selected and read on the small monitor. The observer determines the speed and direction of travel.

Users:

Single user interface.

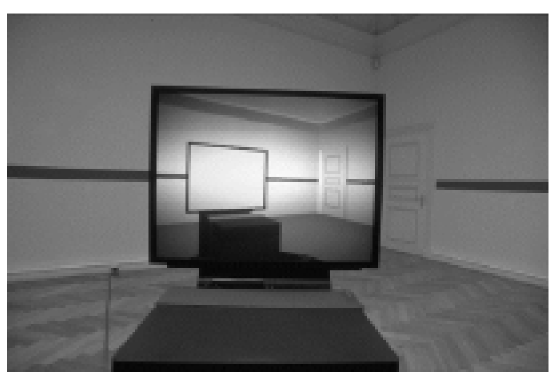
Transform:

The bicycle used as a metaphor for locomotion allows the user to navigate in virtual space in a familiar way. The bicycle as an device is so obvious to use that, from few exceptions, visitors would jump on it and use it right away. It is intuitive and reduces fears in using technology.

Use of Space:

In this installation the users body and with it the notion of body space is integrated in the environment. Sitting on a bicycle and being physically active on the reading journey the user is aware of being situated in the environment: He physically feels himself interacting with the virtual environment.

4.2 The Virtual Museum (Jeffrey Shaw, 1991)



Device:

Armchair on an electronic swiveling platform, motion tracking

Setup:

On a turning platform, there is a chair mounted on front of a rostrum with a superscreen. The observer sits on the chair and can steer the picture on the superscreen by turning the chair and moving his body. The starting sequence offers a mirror-image of the area; the chair is empty. The user can navigate through four museum rooms that show objects of genres of art.

Users:

Single user interface.

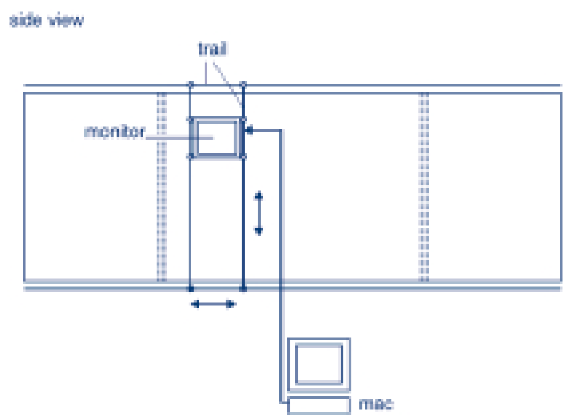
Transform:

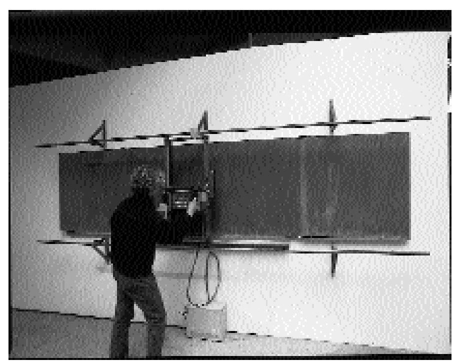
The turning of the platform corresponds to the turning in virtual space giving the user some kind of synchronous alignment with the virtual space and some “force” feedback while traveling. The forward and backward operation, triggered by leaning forward and backward respectively, triggers a corresponding movement in virtual space. It involves the whole body in the navigation while the user still remains in a comfortable rest position in the arm chair.

Use of Space:

Here space is mostly used in an orientational way, i.e. the orientation of the platform towards north or east directions is aligned with the corresponding directions in virtual space.

4.3 Tafel (Frank Fietzek, 1993)





Device:

computer monitor on a carriage in front of a chalkboard.

Setup:

Hanging from the wall is a large chalk board with the side panels folded out. There are traces of smudged chalk on the green surface. In front of the board is a carriage with a monitor. Using two handles, it can be moved up and down or along the carriage sideways. When one searches along the surface of the board by moving the monitor, sentence fragments and single words appear at random.

Users:

Single user interface.

Transform:

A small monitor hangs in front of a bigger chalkboard. Both elements are known as carriers of text and stand for different cultural practices of learning and writing. The presence of the monitor refers to the technological innovations and the fundamental changes in the storage and utilization of information. The installation uses a mobile window paradigm: By moving the monitor, namely pushing and pulling it upwards, downwards, and sideways, the use is actively searching for words and text on the chalkboard that only become visible in the monitor. Here we get a one to one correspondence of the movements in real space to the movements in virtual space (as opposed to a 1-to-3 correspondence of speed in the *Legible City*). By placing the

Monitor in a larger physical environment that directly corresponds to the virtual environment, the interface is both intuitive and does reinforce the storyline. The user perceives a high level of consistency and harmony while using the interface.

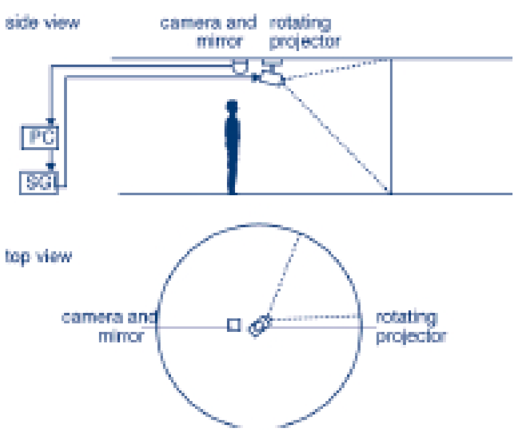
Use of Space:

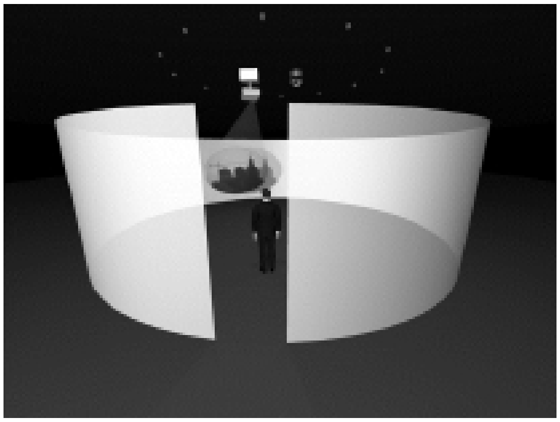
The real chalkboard not only defines the interaction space for the interface itself, but also defines the whole situation the user is placed in: The user is situated in a familiar environment in front of a chalkboard, making it possible to recall memories from childhood. By using the interface, i.e. moving the monitor across the board, the user also moves in real space in front of the chalkboard. This greatly enhances the perception of navigation in the virtual text space in a natural way.

5. Gestural and Unencumbered Tracking Interfaces

In this section gestural interfaces are presented. These gestures need not necessarily be meaningful gestures like handwaving or complicated gestures like showing a combination of fingers. The installations deal with body movements, and body dynamics, in general, they incorporate the expressiveness of human motion in the environment creating an experience that is totally different from manipulating with mouse, for example. An important issue with theses systems is the use of space. As opposed to working with a desktop computer, such systems do need a spatial freedom for the user to express his “gestures”.

5.1 Fugitive (Simon Penny, 1997)





Device:

Video camera based vision tracking system

Setup:

The arena for interaction in Fugitive is a circular space about ten meters in diameter. The infra-red video camera based vision system is used to track a single person in the space, which is invisibly illuminated by infra-red lighting. This camera looks at a reflection of the whole space on a mirrored semi-sphere mounted under the ceiling in the center of the cylindrical space. At the simplest level of interactive feedback, a video projected image travels around the walls in response to a single user’s position. At first, the movements of the image is tightly coupled to the movements of the user, confirming that the system is indeed interactive. However, absolute position of the tracker does not necessarily correspond to a specific location in the virtual space. The image also exhibits other behaviors that correspond to the movements of the visitor’s body over time.

Users:

Single user interface.

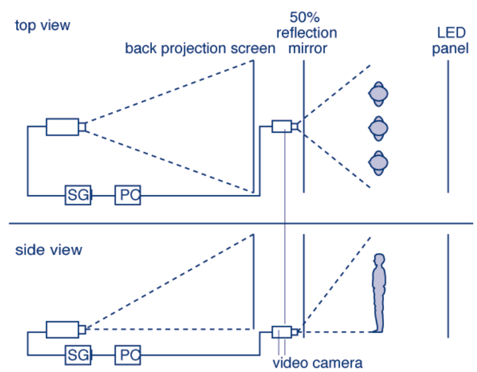
Transform:

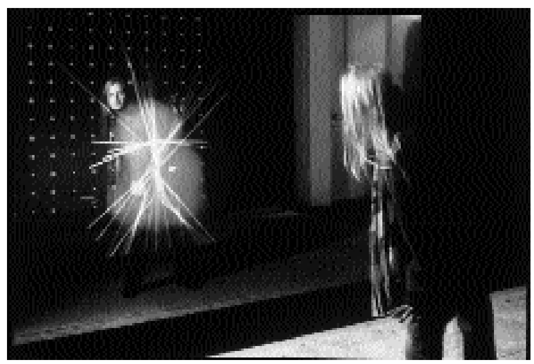
The artist is a well-known theorist of electronic culture, who has criticized simplistic models of interactivity based on positionality. In this installation his formal goal is “to build a system which responds to the bodily dynamics of the user over time, that speaks the language of the body and that is triggered by physiologically meaningful events.” The mapping of body gestures to the flow of digitized video imagery is dependent on the bodily dynamics. “Ideally, changes in the behavior of the system will elicit changes in the user’s behavior, and so an ongoing conversation rather than a calling of Pavlovian responses will emerge.”

Use of Space:

The user as well as the projected image that moves around the circular space is given enough space for expression. An important issue when incorporating gestural input is, to leave the user enough freedom to express those gestures.

5.2 Gravity and Grace (Yasuaki Matsumoto, 1994-95)





Device:

Video camera based vision tracking system, large 50% reflection mirror

Setup:

The observer is confronted with the dark surface of a half-mirrored glass on which blinking red LED lights and the observer’s own image are reflected. As the user enters the scenario, blue shafts of light begin to radiate from around the observer’s image on the glass like an aura. A video-based tracking system is used to make sure the light beams follow the movements of the observer’s body. The software application linked to this vision system also controls specific graphic responses to physical contact between people standing in front of the mirror, and the duration of each person's interaction with the piece.

Users:

Single user interface.

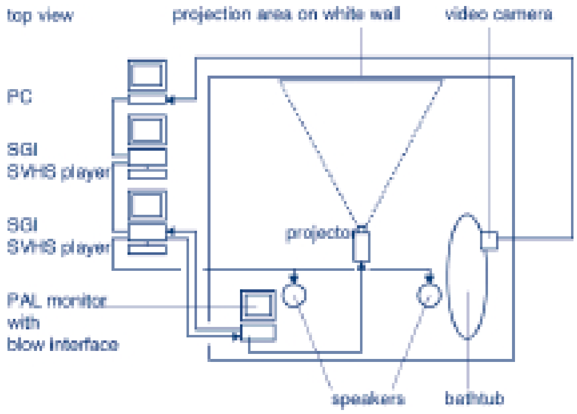
Transform:

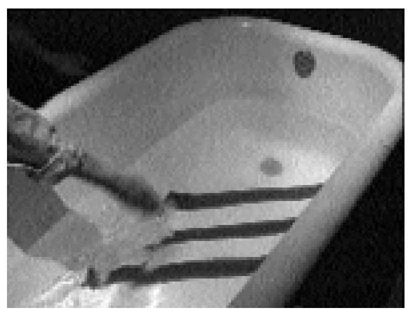
The 50% reflection mirror creates the impression that the users image and the computer generated objects are optically in the same plane and, hence, creates the illusion that graphical objects are being attached to the body. The paradigm used here is similar to the so called “magic mirror” paradigm described in [4]. It allows an easy interaction with a virtual world and leaves the tracking requirements relatively simple. The sensation here is slightly different because here the user is in fact looking in a mirror. He can see himself in a different view, a kind of aura that accompanies him becomes visible.

Use of Space:

Space here is mainly used to emphasize the mirror paradigm which is integrate in real space. The user interacts in real space which, in turn, triggers events and graphics in virtual (the mirror) space. Both spaces are not seamlessly integrated into one single environment which is exactly the same sensation visible when looking into a mirror and therefore need not to be a lack.

5.3 The Wind that Wash the Seas (Chris Dodge, 1994-95)





Device:

custom blow interface, video camera

Setup:

There are two participants involved simultaneously in this installation, the “wind actor” and the “water actor”. On each side of the installation there is an interface for each actor to influence the visual information that is projected onto the back wall. The wind actor blows, lightly or vehemently, against a video monitor. The direction and severity of the gusts are recorded by using heat sensors at the four monitor corners. By means of a large white bath-tub the water actor can interact with the visual environment simultaneously. As soon as the actor's hand churns up the water in the tub, the computer algorithm records the turbulence by calculating the distortion of three black bars that are located on the ground of the tub. Both interfaces are linked to the image-processing programs, and therefore influence the type and extent of image transformation.

Users:

Two users can interact simultaneously.

Transform:

Here somewhat intuitive gestures like blowing and interacting with water are brought to the digital world. The blowing interface here seems to be more direct in a sense that the user is directly blowing onto the image. Whereas, the tub interface is an indirect interface that uses hand eye coordination for controlling the output.

The setup with a real tub filled with water creates a familiar environment with an unfamiliar effect, i.e. the devices used are normally not used in this particular sense. Nevertheless, the link between gesture and reaction is somewhat direct, intuitive, and creates a sensational experience which is completely different from using a mouse, for example, to create turbulences in the image. Here spatial devices are carefully chosen to create this sensation.

6. Conclusion

In this paper, we explored some of the interactive installations at the ZKM in Karlsruhe in five different categories and pointed out some of the basic paradigms of human computer interaction used. We tried to focus on the idiosyncratic use of devices and the transformation of these within the context of the work. It is specially significant to note that the success of these input and tracking devices is largely due to the careful manner in which the artists have chosen interface strategies that are exactly appropriate to the specific content of their works, even in those cases where common generic devices were used. This cohesive and consolidated design approach to the interactive form and content of each work also guarantees the intuitive transparency and ease of use of these interfaces, even in those cases where the user is confronted by very unusual situations. We conclude that a complete spectrum of interface environments, ranging from familiar to innovative, from simple to complex, from mechanistic to unencumbered, can be successfully exploited in eRENA applications, as long as in the choices made there is a harmonious correlation between their functionality and the content articulation of a specific application.

References

[1] C. Blase, M. Morse, A. Blunck, et. al. (1997), *“Hardware, Software, Artware : Confluence of Art and Technology. Art Practice at the ZKM Institute for Visual Media 1992 –1997"*”, Cantz Verlag, 200 pages + CD-ROM, ISBN 3-89322-928-0, 30,- DM.

[2] H.P. Schwarz (1997), *“Media —Art —History”*, published on the occasion of the opening of the Media Museum, ZKM | Center for Art and Media Karlsruhe on October 18, 1997, Prestel-Verlag, Munich, 1997. 191 pages + CD-ROM, ISBN 3-7913-1878-0 (english version)

[3] J. Foley, A. van Dam, S. Feiner, J. Hughes (1991), *Computer Graphics. Principles and Practice. Second Edition.* Addison-Wesley, 1991.

[4] T. Darrell, et. al. (1994), *A Novel Environment for Situated Vision and Behavior*, Proc. of CVPR-94 Workshop for Visual Behaviors, pp. 68-72, Seattle, Washington, June 1994.

Part II
Navigation for the Senses

Jasminko Novak, Monika Fleischmann, Wolfgang Strauss,
Thomas Kulessa

1. Introduction

Navigation depends on the functionality of the interface. Metaphors of navigation are building the tools for it and vice versa. The first step in this task is to create an interface environment connecting inter-actors to the system. The sensation of the body will transform a given virtual environment into a field of emotion. Instinctive interfaces will support several senses of the body. The level of transformation and deconstruction depends on a variety of matters: distance and approach, speed of movement, skin temperature, gesture based tracking, camera tracking as a series of parameters to influence the virtual surrounding. The goal is to realize a virtual chamber of awareness and sensitivity to real people in order to develop a natural relation to rigid I/O systems. Elements of the interaction like body balance, body movement or gesture expression are basic elements of human performance in space. As a body centered platform the Virtual Balance system will be discussed. Conceptually the Virtual Balance will be further developed as an input and navigation device for two users at different places. As a result of evaluation of the Virtual Balance a catalog of features was worked out. Major technical solutions could be developed and implemented in the second year of eRENA.

2. The Metaphor of Navigation

The term "navigation" signifies the definition of and adherence to a course and is derived from the Latin "navigare" which can be translated as steering, sailing or travelling. The same symbols are used in virtual space as in real space—though virtual navigation involves the "re-configuring"—i.e. production—of a time process.

The voyages of discovery made in the late Middle Ages radically changed the geographical nature of the world. The records made by navigators provided new information on the number and location of the continents. European philosophy was shaped by the travel reports of the 18th and 19th centuries made by such persons as Charles Darwin. They stimulated debate on the possible diversity and relativity of thought. An evaluation of the new findings provided a basis for addressing the shortcomings of one's own society and of formulating new state theories.

Etymological dictionaries define the term expedition, as a voyage of research. The term was derived in the 16th century from "expedire" which can be translated as "unshackling one's feet". Marcel Duchamp coined the phrase "My feet are my studio" in the 1920s and saw this "liberation from the shackles" as an instrument with which one could learn to recognize and understand space.

2.1 Space and Communication

The concept of space in the 20th century has changed from the idea of conquering space to one of its dissolution and has been brought about primarily through the new means of transport which have become available.

After the Hubble-Space Telescope was launched in 1990, from 1993 on the NASA camera was able to send spectacular pictures received from the depths of the universe. "Hubble" has allowed us to see further and more "clearly".

The travelling tradesman of old had a communicative function. On his travels through the world, he acquired information and passed this on to other persons he encountered. The troubadour, too, transformed the information he had acquired into the form of songs. The "Dissidents"—a group of German musicians—is today devoted to creating informative and communicative "World music" by teaming up with local musicians as they travel through the world.

2.2. Virtual Space and Navigation

Is this "culture of interactivity" also possible in virtual space?
Can the same metaphors which are used for exploration and orientation in real space also be applied to describe virtual space? And what does orientation in virtual space actually mean? In virtual space we practice for reality and live with a feeling of 'as if'. We simulate and practice communication processes.

The concept of interactivity is generally limited to the simple "selection" of information. The navigation concept could, however, devise mechanisms for links for making virtual space tangible. The idea of electronic arenas is to create networked virtual space to build virtual communities. The computer platform and the use of a specific software constitutes only part of the work in search of orientation. But, far more important are the own thoughts and the process involved—what groups of people can communicate better using electronic arenas?

The expeditions to the virtual world are bodiless. Nevertheless, the body has not disappeared. The link between body and virtual navigation space is often hindered by keyboard and mouse. Navigation in imaginary virtual spaces requires interfaces which allow the participant to travel between the various worlds in order to create an illusion space. Rediscovery of the senses leads to methods for developing poetic interfaces which give us a new sense of the senses. [1]

3. The "Virtual Balance"—looking with the feet.

3.1 Navigation through body balance

The *Virtual Balance* was developed at GMD in 1995 [24]. Like Hermes the celestial messenger, the observer navigates through a digital landscape by using "virtual balance". To do this, he simply has to move his body's centre of gravity to allow him to fly upwards or downwards, to the right or to the left. The dramatic effect of the action is governed by the person's relationship to his own body. Here, we observe physically handicapped persons who are motivated in their movements. The ground below their feet becomes an interactive surface and the body's perceptual sensitivity coupled with body balance becomes a control instrument.

Unlike joystick and mouse which reduce the human being to a small set of reflexes [22], the *Virtual Balance* requires the use of the whole body and perception. The *Virtual Balance* (VB) is a performer-centered reaction device with sensors connected to an interactive virtual environment. As in the real world, the body becomes the control tool of the virtual environment. Technology is going to be like life, the virtual balance requires real balance. The *"Virtual*

Balance" interface is based on man-machine interaction by movements of the human body on a sensored platform. Thus, by shifting of weight or minor movements the actor controls his position in a virtual environment.

The *"Virtual Balance"* consists of a platform with three force sensors and is controlled solely by the changes in the position of the human body's center of gravity. The observer's positional information is passed to the graphical system for the purpose of calculating the image for the current viewpoint. At the same time it is also a platform for observing the effect of images on the body. During the presentation at CeBit '96 in Hanover, neurologist Hinderk Emrich found himself repeatedly in dance situations and discovered an "enthralling" perspective of the virtual world.

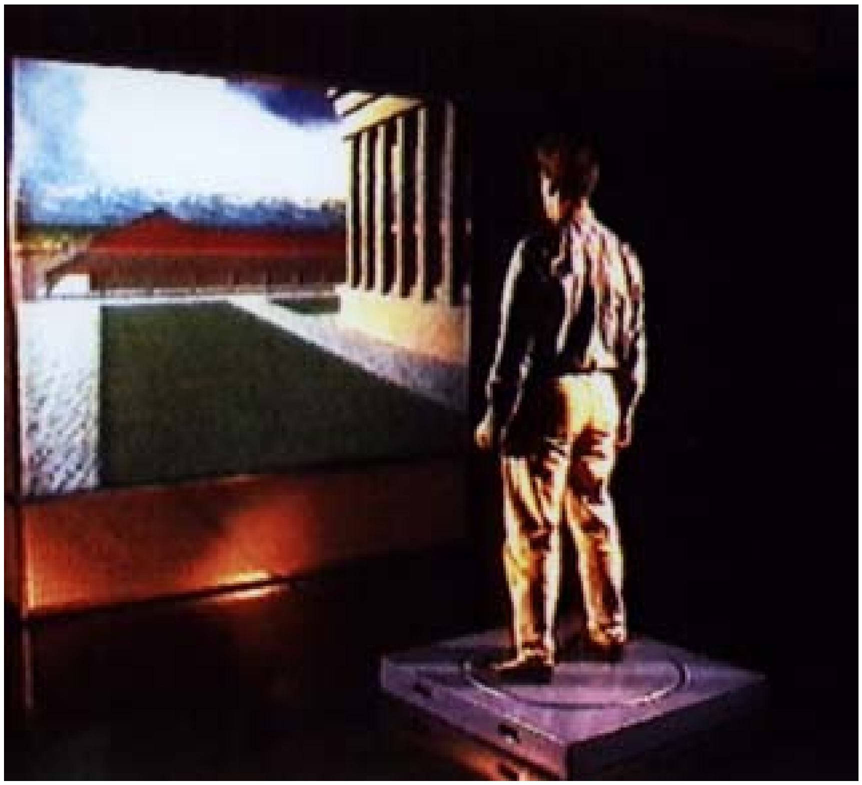


Fig. 1: Navigation with the Virtual Balance

3.2 Evaluation of the Virtual Balance

We have evaluated the Balance in a walk-mode where the weight shifting causes motion in the horizontal plane only, since height is not needed for navigation in closed spaces.

Advantages of the Virtual Balance are:

- hands are free for other tasks, such as using a theremin interface as outlined in section 4.3.
- navigation requires no effort—albeit this can also be a disadvantage for the feeling of immersion.

Disadvantages:

- There is little precision of navigation
- It is likely that equilibrium problems will occur for users in an HMD (due to simultaneous orientation in space through head movements and body movements for the Virtual Balance)
- The small sample rate of the system makes navigation difficult and the problem is

higher sample rate would be needed for a real evaluation of the Virtual Balance. This would require a new A/D converter and new driver software.

One could try out a few more things with the software model:

- As the VB is an isometric (nonflexible) device, it is best utilised as a speed control device [22]. (Leaning forwards and backwards causing speeding up and slowing down.) Addition of some amount of elasticity, if this can be done without sacrificing the ruggedness of the device, will likely lower the learning threshold of the device [22].
- As it is difficult to stand exactly still, one should define a neutral zone where small deviations from the home position are ignored [20].
- For moving in the vertical dimension one could incorporate “escalators” or “lifts” in the virtual environment.

As an extension of the Virtual Balance we would like to introduce a concept of a space related interface able to measure the position of a body in space, possibly through video tracking as described in section 5.2 of this document. This would enable simultaneous navigation and manipulation.

3.3 Virtual Balance in a connected navigation system

After evaluating the approach of connecting two virtual balances directly we have decided to take a more flexible approach which actually allows connecting several Virtual Balances instead of just two.

The solution is to connect the balance to a VRML browser where it can serve not only as a navigation device but also as an interface for special scenarios. VRML was chosen because it is a standardized networked scene description language which (coupled with Java for dynamic scene modifications) enables us to create different settings in which the balances could be used as means of interaction and communication between participants, and not merely for navigation in the scene.

GMD’s FIRST Institute in Berlin has developed a driver for attaching the Virtual Balance to a VRML browser and they have tested it with the VRWeb browser. The driver takes the output of the Virtual Balance and converts it into data appropriate for controling the VRML browser.

We developed the MARS External User Interface Driver which enables connecting any desired input device to any VRML browser supporting the External Authoring Interface [18] for Java-VRML communication (such as CosmoPlayer). The External User Interface Driver is described in more detail in Task 1.1 since it was used there to interface the MARS optical tracking system to control movement of avatars in a VRML scene.

The next step would be to construct another Virtual Balance in order to be able to experiment with possible scenarios involving two balances connected through the described connected navigation system for several balances.

The Virtual Balance as a navigation tool for a VRML browser will be publicly demonstrated on June 26–28, at the Performance Symposium in Potsdam, Germany.

4. Interfacing the theremin

Developing the possibility of using the theremin as an interface illustrates developing a new paradigm in human-machine interaction. This is because the theremin by its nature "communicates" directly with the human body and its properties as physical matter. Electromagnetic tracking systems use special devices carried by people to measure the strength space. Optical interfaces analyze an image, treating it as a set of abstract pixels, where the only information is of colors or shapes which have nothing to do with the human body.

In contrast, the theremin reacts directly on the physical condition and properties of the human body such as capacitance and conductance, thus metaphorically exemplifying the attempt to explore new ways of reading human bodies with computer systems, starting with their physical properties. When built into a beautifully designed wooden housing a theremin interface is a haptic toy with an imprint of the implemented functions.

The goal of this work is to connect the theremin to a computer system as a simple movement and gesture interface, to serve as input for performances in virtual environments.

4.1 Principles of the theremin.

The Interface "Virtual Theremin" is based on the "theremin", an invention of the Russian physicist Leon Theremin (Lev Termen) in 1919 [25]. The theremin was one of the first electronic musical instruments.

The theremin is played by waving one's hands near two metal antennas: one for pitch and the other for volume. To create the sound, a fixed oscillator is mixed with the variable pitch oscillator and their difference (or beat frequency) is amplified.

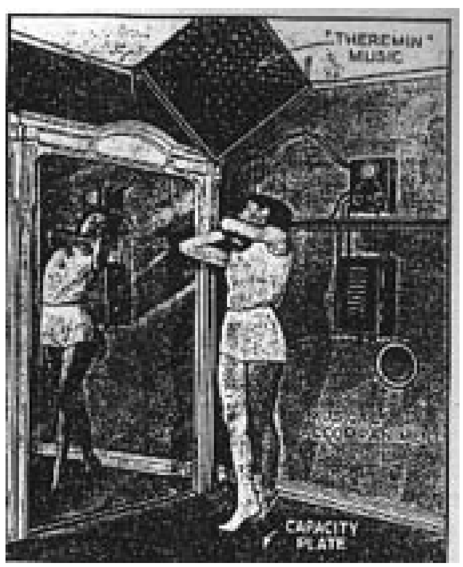


Fig.2: The Terpsitone (*Radiocraft* Dec. 1936, p. 365)

With Theremin's "Terpsitone", which depends, like the theremin, on the capacitance of the body, it is possible for a dancer to dance in tune and in time. In place of the rods used in the first theremin there is an insulated metal plate beneath the dancing floor. As the dancer bends towards it, the electrical capacitance is increased and thereby the pitch of an oscillating tube circuit is lowered; if she, for instance, rises on tiptoe, the pitch of the oscillator is increased. Thus the motions of the dancer are converted into tones varying in exact synchrony with her pose. In fact, the motion of either one arm or a leg is sufficient to produce a noticeable change of tone. In the "Terpsitone" configuration the loudspeaker used to give this individual tone interpretation of the dance is supplemented by another, reproducing a background of the theme music previously selected.

As the theremin is based on the capacitance of the body parts close to the antennæ, playing can be said to be based on “body mapping”—no single point on the body is controlling the sound, but instead any part can affect the sound. Musicians playing melodies seem to perceive the space around the theremin as a haptic space or a “virtual screen” on which they “feel the touch” of the correct tones. This demonstrates the very tight coupling between hand movement and sound perception. While having a very sharp control over the sound, one can also slide through

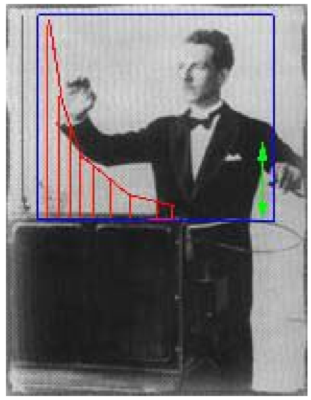


Fig.3: The virtual screen of a theremin.

We see "Virtual Theremin" applications in fields where fast no-touch controls are essential, as well as in the area of performance. A theremin interface could be used in outdoor areas, in public spaces, but also in workspaces. It could easily be built weatherproofed (therespace). A two-sensored theremin could be used as a desktop computer interface (thereface). We see the "Virtual Theremin" as a teletouch interface with untethered gestures to Navigate/manipulate/generate data like images, sounds or space. Therefore the most interesting idea is to use several theremin interfaces for artistic performances.

We think that the theremin could be used as a

- gestural navigation interface;
- instinctive, unnoticed interface;
- outdoor interface;
- tracking interface.

4.3. Theremin as a device for gestural navigation

We have explored using the theremin as an input device in a twofold manner:

- as a navigational input device for VRML environments,
- as an unsharp gestural interface device.

The idea of using the theremin as an unsharp gestural interface device is that it reacts to user gestures over time instead of responding to precise, pre-defined, command-like movements. We found that the theremin signal is not suitable for such a control scheme because it carries such a small amount of information (one dominant frequency followed by low amplitude harmonics replicating the behaviour of the dominant). In the tested configurations, one theremin is best used to control movement in one direction.

A standard theremin has two antennæ (but cf e g [23]) and we thus have the same problem as with the Virtual Balance to map the available degrees of freedom to movement in 3D space. The volume control (left hand) values could be used to control speed of movement, the pitch control (right hand) to indicate movements right or left as in the walk-mode in a VRML browser. Switching to fly-mode could be done by touching the pitch antenna and thus give an easily recognizable signal. In flying mode movement up and down could be controlled by the left hand, while turns are indicated with the right hand. An example application is using the theremin in conjunction with an optical tracking system to track users' gestures and movements in the demonstrator "Murmuring Fields" in Task 1.1.

Extending the theremin with multiple antennæ gives us additional degrees of freedom, and each hand could conceivable control three linearly independent sensors. An interesting result is that if configured as a theremin floor (see section 4.4), the dimension of distance, coupled with the detection of intensity of presence (number of users in its reach) is enough for simple mechanisms of viewpoint control. Increasing the number of elements provides more sophisticated possibilites.

Theremin output data is captured via a common audio port and interpreted using our MARS Simple Gesture Interface Driver. This driver implements several schemes for interpreting raw audio data but a simple frequency scan and a sliding-window Fourier transformation has proved most usable at present. The first provides a very good spherical-distance measure and is suitable when using different theremin configurations as simple navigational devices. The second is better when the theremin is used as an unsharp interface or when only influencing an existing movement to produce slight modifications in it, rather than exerting total control over it.

4.4. Theremin floor as invisible interface for CAVEs and other VEs

We have developed the concept of a theremin floor as an invisible and intuitive interface suitable for virtual environments such as a CAVE or artistic performances in hybrid space environments (such as the "Murmuring Fields" installation in Task 1.1).

The CAVE seems a natural starting point because this approach solves the problem of only one user actively navigating in the scene. A sensitive floor could be used to enable multiple users to influence navigation inside a CAVE.

This device is being realized using a number of theremins equipped with a flat pitch antenna covered by the floor. The theremins divide the floor into patches where the presence or non-presence of users is localized. The number and position of users within a patch influences the theremin output signal which provides the information to be interpreted for navigation in the scene.

GMD is realizing the theremin floor for the extended performance with audience participation at the Cyberstar award ceremony on June 14, in KOMED Center in Cologne, Germany. On this occasion we will also produce the video for the demonstrator.

The current solution for the connection of the theremin floor to the CAVE builds on GMD’s modular architecture for connecting external input devices to VRML based virtual environments. Theremin output data are processed and interpreted by the MARS Simple Gesture Interface Driver whose output is passed to the MARS External User Interface Driver. The navigation in the 3D scene is done by connecting this data stream to our Simple Shared Environment Server which directly controls the VRML scene in the browser. Accordingly, browser output is projected into the CAVE. What would need to be resolved further is stereo projection, which is not supported by the browser. A simple workaround could be using one user data stream to simultaneously control two slightly displaced viewpoints of the scene.

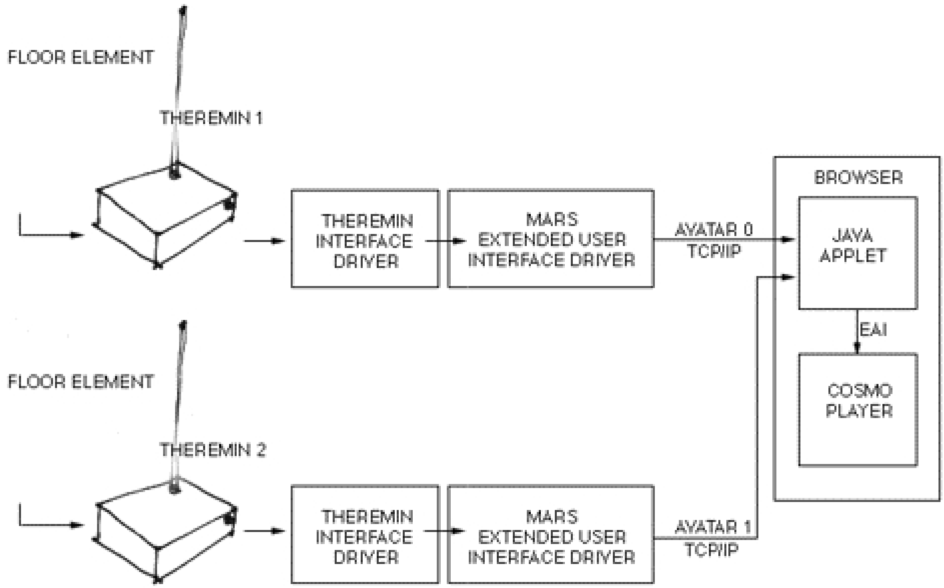


Figure 4. Simple two-patch theremin floor connected to the VRML browser.

This implementation is in a very experimental phase, and a lot more work would need to be done to ensure stable functioning of the theremin floor due to the very sensitive behaviour of theremins with regards to mutual interference, as well as environmental conditions (temperature, proximity of metal objects etc.).

Another problem that needs to be resolved is finding a good scheme for coordination of simultaneous influences of several users to the viewpoint control, i.e. a viewpoint control protocol.

5. Camera based person tracking

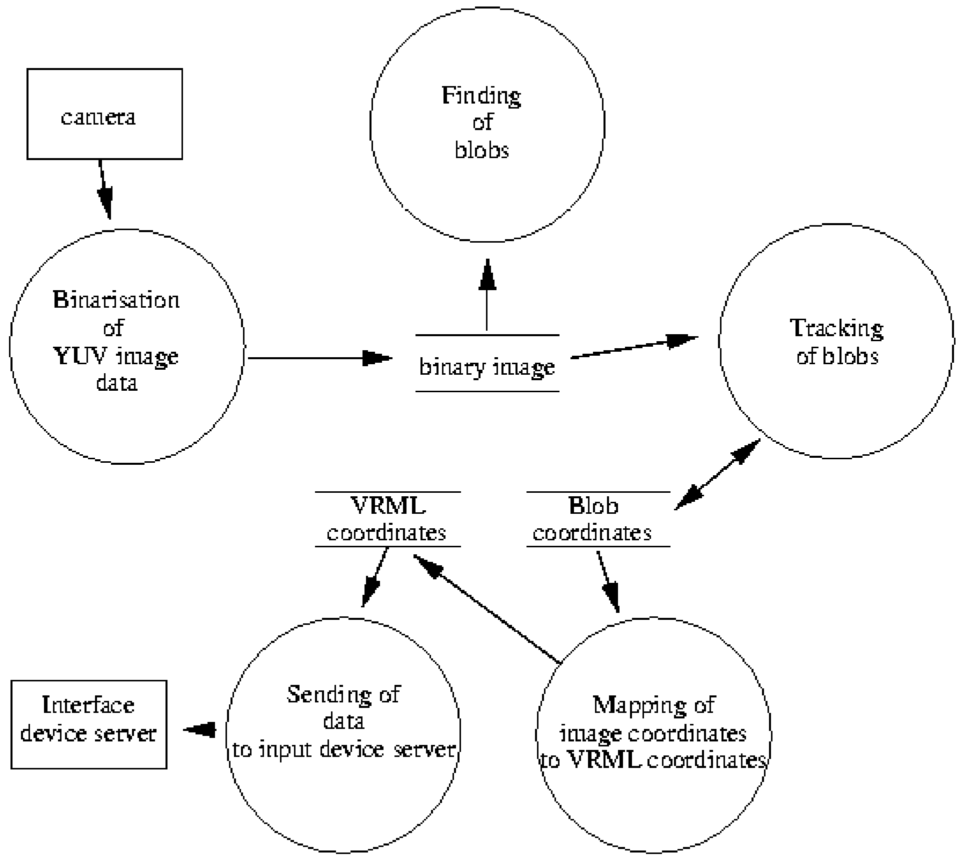
A straightforward approach for avatar navigation in virtual scenes is to map the body movement of an observer of a scene directly onto the movement of the avatar. One way to accomplish this is through the use of an *camera based human tracking sytem*. Such systems try to find humans in images taken from a camera observing a real space using form or color information.

For the use in interactive media art installations such a system has to satisfy the following requirements:

- It has to work in *real-time*, which means that the avatar in the virtual scene responds with no or at least very little time delay to the movement of the observer in the real scene.
- For use in public installations it is necessary that the camera tracking system is *independent of the appearance* of the participants.
- To give the media artist the freedom to design the observed space in a fashion he likes, it is desireable that the camera interface works in different scenarios with no constraints on *illumination conditions* and *spatial arrangement* of the objects in the observed scene.

Since there is a broad spectrum of possible areas to use cameras as interface to a computer, like psychology, intelligent home enviroments and film-planning, much work has been done on this topic [2, 3, 5, 13, 14, 15, 16].

In this section we will describe the person tracking scheme used in the works on the "Extended Home of the Brain" and "Murmuring Fields". First some related work, which is part of the



techniques used for body tracking are depicted. Finally, we take a look on possible further development that can be done to improve the system.

5.1. Related work

As mentioned above much work has been done on camera tracking. One of the earliest applications for artistic purposes is Myron Krueger's VIDEOPLACE [26], which later has been followed by the Mandala system from the Vivid Group. Hoch, in his work on the "Intuitive Interface" [6, 8] developed the C++ class library mTRACK which defines a set of functions to interface a virtual scene in real-time through vision and speech based devices.

In this work we use this framework to implement our own techniques for *image segmentation*. In [12] a color calibration tool using mTRACK was developed. It is used in this work to determine the needed starting parameters. The next subsection describes what these parameters are and how they are acquired.

5.2 Description of camera tracking

Fig. 5: Data flow of human body tracking system.

As can be seen the process of tracking is divided into five steps:

- The *binarization of the input image*, which means the division of the input image in areas which belongs to the background and areas which could be a person.

- Once this division is made, the set of non-background pixels has to be further analysed to separate individual persons in the observed scene. This process is called *blob analysis*.
- After the blobs in the input image representing a person are determined, these regions have to be *tracked* in the subsequent images.
- In the next step the two-dimensional image coordinates of the tracked regions have to be *mapped* on 3D-VRML coordinates to allow navigation in a virtual scene.
- Finally, the transformed coordinates are send to an *interface device server*, described in an other section of this paper, which forwards them to the connected VRML-Java applets.

The following paragraphs will describe each step in more detail.

Basically there are two ways to binarize a given image. One way is to find regions in the image resembling the human shape. This is a very time-consuming process and on input data of low quality nearly impossible. For this reason most real-time tracking systems use an other approach. They try to binarize, or *segment*, the input image just by color information or use a mixed approach for analysing spatial and color information [2, 3].

In this work we use a very simple but efficient scheme for color segmentation, namely segmentation through *thresholding* the image. Schroeter shows in his work [16, 17], that the YUV color space is suitable for color segmentation under varying illumination conditions. It can be calculated linearly from RGB data but has the advantage that it separates brightness (Y) from color (UV) information which makes it easier to determine the "color" of a person.

For each image pixel it can be determined if it is part of the background or the person by checking if the U and V values fall within given intervals. The threshold intervals are determined interactively with the help of the tool *multigrab* developed by Kulessa [12]. The method of image segmentation by thresholding only produces good results in scenes with no colors in the background similar to that of the tracked persons. At the current stage of our work we do experiments in a blue room with persons dressed in black. In this scenario this approach is totally sufficient. The next section describes a better approach to overcome these shortcomings, which will be implemented in the future.

The next step in the process of body tracking is the determination of connected regions in the input image. If two persons in the observed scene are not standing too close together each region represents a human in the scene. Small regions with less than forty pixels are discarded, because they are assumed to be effects of image noise or part of the background. The determination of connected regions is done by an *8-connected*-neighbourhood analysis. This is done by mTRACK through the use of the Matrox-MIL-library and will not be further described. A detailed description can be found in [5].

After the connected regions have been determined, these regions have to be tracked through the following images. This is done by the use of a *Kalman filter*. This filter uses information about the position of the tracked object in the past to predict the position of the object in the current image frame. Details about this can be found in [10, 15]. One reason the Kalman filter is used, is to distinguish each single person in the scene, another is to reduce the amount of image data to be processed by examining only the image regions in which the tracked objects are most likely to be.

Once the (x,y)-positions of the tracked persons are acquired, these coordinates have to be transformed to use them as input for the navigation of an avatar in the virtual scene. In a first step we map the (x,y)-coordinates on the (y,z)-coordinates of the VRML scene. Through this it is possible to move the avatar in a plane in the virtual space. The z-coordinate can be manipulated by the theremin as described above. The next section describes an improvement to get real 3D coordinates to make the avatar navigation more intuitive.

The last step in the tracking process is to send the data to the VRML scene. This is done by writing the data to the console, where it is read by a server which sends them through a socket connection to different Java applets.

The system was implemented in C++ on standard PC hardware. In runs at a speed of 10–20 frames per second.

5.3 Future work

We have implemented a system, tracking a number of persons in a camera observed scene. This is done in real-time. We described a simple, but efficient method for image segmentation.

There are two major improvements that could be implemented in the future:

- the use of a second camera to gain real 3D information.
- the use of a more sophisticated method for image segmentation.

The system has to somehow identify the 2D coordinates of a point on the body of the tracked person. A good approximation of this is the center of gravity of the segmented blob [15]. The second major improvement is the use of color *look up tables* (LUTs). They have two advantages over the segmentation by thresholds. They are more robust to similarly-colored objects in the background and they produce better tracking results under changing and inhomogenous lighting conditions. Approaches to calculate them are described in [4, 13, 17].

References

[1] Monika Fleischmann, and Wolfgang Strauss. "Images of the Body in the House of Illusion". In C. Sommerer (ed.), *Art@Science*. Vienna/New York: Springer, 1997.

[2] Ali Azarbayejani and Alex Pentland. "Real-time self-calibrating stereo person tracking using 3-D shape estimation from blob features", *Proceedings ICPR '96*, 1996. `ftp://whitechapel.media.mit.edu/pub/tech-reports/TR-363.ps.Z`

[3] Ali Azarbayejani, Christopher Wren and Alex Pentland. "Real-Time Tracking of the Human Body", *Proc. of IMAGE'COM 96*, France, 1996. `ftp://whitechapel.media.mit.edu/pub/tech-reports/TR-374.ps.Z`

[4] Gary Bente. *Computersimulation nonverbalen Interaktionsverhaltens*, 37. Kongreß der Deutschen Gesellschaft für Psychologie, Kiel, 1990

[5] J.L. Crowley and J. Coutaz. "Vision for man machine interaction", *Proc. of EHCI '95*, Grand Targhee, 1995

[6] Michael Hoch. "Social Enviroment: Towards an Intuitive User Interface", *Proceedings 3D Image Analysis and Synthesis '96*, Erlangen, 1996

[7] Michael Hoch. "Object Oriented Design of the Intuitive Interface", *3D Image Analysis and Synthesis: Proceedings*, Erlangen, pp 161–167, Infix, November 17–19, 1997.

[8] Michael Hoch. „Intuitive Schnittstelle", *Lab, Jahrbuch 1996/97 für Künste und Apparate*, Verlag Walther König, Köln, 1997

[9] Michael Hoch. "A Prototype System for Intuitive Film Planning", *Third IEEE International Conference on Automatic Face and Gesture Recognition FG'98*, April 98, Nara, Japan.

[10] Markus Kohler. *Using the Kalman Filter to track Human Interactive Motion— Modelling and Initialization of the Kalman Filter for Translational Motion*, Technical Report 629, Univ. Dortmund, 1997

[11] Markus Kohler. *Technical Details and Ergonomical Aspects of Gesture Recognition applied in Intelligent Home Enviroments*, Technical Report 638, Univ. Dortmund, 1997

[12] Thomas Kulessa. *Automatische Kalibrierung zur Personenverfolgung in Farbbildfolgen*, Diplomarbeit, Univ. Dortmund, 1998

[13] Thomas Kulessa and Michael Hoch "Efficient color segmentation under varying illumination conditions", to appear in *IMDSP '98*, Alpbach, 1998

[14] Projektgruppe 292. *A-Help — Optischer Sensor zur Erkennung von medizinischen Notfällen im Wohnbereich*, Technical Report, Univ. Dortmund, 1998

[15] Projektgruppe 277. *Argus — Ein ergonomisches System zur Steuerung von Haushaltgeräten mittels Gestenerkennung*, Technical Report, Univ. Dortmund, 1997

[16] Sven Schroeter. *Entwicklung von Verfahren zur automatischen Kalibration eines farbbasierten Objekterkennungssystems*, Diplomarbeit, Univ. Dortmund, 1996

[17] Sven Schroeter. "Automatic Calibration of Lookup-Tables for Color Image Segmentation", *3D Image Analysis and Synthesis (Proceedings)*, pp 123–129, Infix 1997.

[18] Chris Marrin. "VRML 2.0 Proposal: External Authoring Interface Reference", 1997, `http://cosmosoftware.com/developer/moving-worlds/spec/ExternalInterface.html`

[19] Bernie Roehl, Justin Couch, Cindy Reed-Ballreich, Tim Rohaly and Geoff Brown. *Late Night VRML 2.0 with Java*, Ziff-Davis Press, Emeryville, CA, 1997

[20] J. D. Foley, A. van Dam, S. K. Feiner and J. F. Hughes. *Computer graphics: principles and pratice*, 2nd ed., Addison-Wesley, 1992.

[21] Joseph D.Rutledge and Ted Selker. "Force-to-motion functions for pointing". In D. Diaper, D. Gilmore, G. Cockton and B. Shackel (eds), *INTERACT '90*, pp 701–706, 1990.

[21] William Buxton and Brad A. Myers. "A study in two-handed input". In *CHI '86*, pp 321–326. SIGCHI, April 1986.
`http://www.dgp.toronto.edu/OTP/papers/bill.buxton/2Hands.html`

[22] Shumin Zhai. *Human Performance in Six Degree of Freedom Input Control*. Ph.D. thesis, University of Toronto, 1995.
`http://vered.rose.toronto.edu/people/shumin_dir/papers/PhD_Thesis/top_page.html`

[23] PAiA Electronics Inc. "PAiA: George McDonald".
`http://vww.qns.com/paia/georgemc.htm`

[24] Monika Fleischmann, Thomas Sikora, Wolfgang Heiden, Wolfgang Strauss, Karsten Sikora and Josef Speier. "Virtual balance: an input device for VR environments".*La Lettre de l'IA*, 123, pp 20–23, 1997.
`http://viswiz.gmd.de/~heiden/wolfweb/publ/virbal/ppframe.htm`

[25] Jason B. Barile. "The Theremin Home Page". April, 1998.
`http://vww.Nashville.Net/~theremin/`

[26] Myron W. Krueger. *Artificial Reality II.*. 1991.

Part III
Some Elementary Gestural Techniques for Real-Time Interaction in Artistic Performances

John Bowers, Sten-Olof Hellström

Introduction and Background

This short chapter documents some of the early prototypes that we have been investigating at KTH concerned with elementary gesture recognition and processing techniques. Our approach is wilfully simple—yet hopefully not simplistic. We have been using relatively inexpensive sensor equipment and analysing the data from various simple configurations of sensors is very straightforward ways. Our approach is resolutely 'bottom up'. In contrast to much of the literature on gesture processing, we perform very few computations on the raw sensor data to identify elementary gestures. Once a gesture is recognised, context sensitive further processing of the data can take place with the identified gesture acting as a 'context switch'. Although we are yet to employ our gesture recognition and processing techniques in a live performance situation, we have some preliminary indication that our 'simplest', 'bottom up' approach coupled to the algorithmically mediated 'expansion' of elementary gesture may yield some promising interaction tools for artistic use. This chapter describes some of the background to our work, indicates its current status through two demonstrations (one involving proximity sensors, the other a simple dataglove) and suggests some future possibilities. Throughout it must be kept in mind that this work is at an early stage of development.

Experiences with *Lightwork*

Deliverable 2.2 of eRENA this year describes an interactive performance work called *Lightwork* developed and performed by workers at KTH. In this piece, navigable virtual environments are constructed algorithmically on the fly while electroacoustic music is performed live. The piece emphasises the improvisation of sound and virtual environment content and a number of interaction techniques were developed for it. In particular, we worked under the auspices of what we call 'algorithmically mediated interaction' or 'indirect manipulation'. Performer gesture does not influence sound or vision directly. Rather, data from performers undergoes a number of transformations before effects are felt in the virtual graphical or sonic environment. In short, performers interact with algorithms, data captured from their interaction devices parameterising the algorithms, quite possibly after further transformations.

In *Lightwork* the performers' interaction devices to date have been fairly conventional MIDI transmitting musical tools of various sorts: a Yamaha WX-11 wind controller, a Yamaha MFC-10 footswitch board, a Peavey PC1600 bank of 16 MIDI faders and so forth. What happens to MIDI data from these devices may be unconventional but the devices themselves are quite familiar. We wished to explore more experimental controllers—particularly for the control of sound in *Lightwork*—but we set ourselves an aggressive target date for a first performance and this meant that some of our ambitions had to be tempered.

In particular our aesthetic concern with making the assembly of virtual environments and soundscapes a 'lightweight' affair requiring a lightness of touch and gesture was not reflected in much of the conventional equipment we used (on these matters see the section on 'Some Aesthetic Themes' in the chapter on *Lightwork* in Deliverable 2.2). The manipulation of a bank of MIDI faders hardly makes for interesting viewing on the part of an audience and, as a material device, it is not aesthetically consistent with the themes we wished to explore. The performer devoted to the processing of the sound (whom we call S in Deliverable 2.2) expressed some frustration in the limits placed on the repertoire of performance gestures available to him—even though, in a technical sense, much was made of the MIDI data he generated (complex non-linear transformations and the like).

For these reasons, we have devoted some preliminary effort to exploring techniques for the processing of sensor data from *non-contact devices* as these seem idiomatic for the aesthetic themes of *Lightwork* and make for an interesting challenge for a musician more used to hands on contact with instruments and other musical devices (cf. also Deliverable 2.2 on 'paradoxical interaction devices').

Expressivity and Gestural Legibility

The other performer of *Lightwork* (whom we call V in Deliverable 2.2) also had some frustrations with the performance environment created for him. These are documented in Deliverable 2.2 and concern the 'downside' of the indirect interaction techniques which otherwise would seem to have an interesting flexibility. In *Lightwork* V's playing of the Yamaha WX-11 wind instrument is analysed in various moving event 'windows'. Statistics are computed for his playing characteristics in these windows and it is these statistics which are transmitted on demand as parameter values (after some further scaling and extrapolation) to the algorithms which generate virtual environment content. That is, it is sets of notes played, rather than any single one, which have an influence over the algorithms used for constructing virtual environments in the piece. Another example of our interest in indirect manipulation. In principle, such techniques should allow a performer some flexibility in the design of expression. For example, risks could be taken as errors can be compensated for within the relevant event window. With further rehearsal and more careful calibration of the technologies we were using, this—we feel—would probably have been fulfilled (and should be in the future). However, as discussed in Deliverables 2.2 and 2.3, a problem with indirect techniques is that they may lead to a rather opaque experience for the audience, who may find it hard to follow just how these 'indirect manipulations' influence what is being experienced.

In *Lightwork*, new virtual content is computed only when a relevant footswitch (to select the appropriate algorithm and signal the correct moment in the improvisation) is pressed. In this respect, V's interaction techniques have a direct manipulation (DM) component. Indeed, V found himself exaggerating his gestures with the footswitches so as to convey his 'live' connection to the projected virtual world. This is slightly ironic, though, as it is the idea of V breathing virtual environments into existence which is an important aesthetic theme for us in *Lightwork*. V should not need to 'stomp' worlds into existence to satisfy his expressivity in performance and yield legible gestures for the audience.

Simple, Loosely Coupled, Hybrid, Gesture Mediated Interaction Techniques

This discussion of our experience with *Lightwork* suggests a possibility worth exploring for supporting performers of artistic work with novel interaction techniques. We feel that our idea to 'loosely couple' performer gesture to any technical system is an important one. It should not necessarily be the case that the slightest quiver could always potentially lead to undesired outcomes. Our performance processing technique in *Lightwork* of concatenating multiple gestures in moving time windows addresses some aspects of this. We also, as performers, feel uncomfortable with the image of the performer's body wired up to multiple sensors, each of sensors (a literal conjoining) would be a straitjacket for many kinds of performance. Accordingly, we prefer sensor architectures where a relatively small number of sensors are available and which can engaged with or disengaged with at will. Unless the close and intimate coupling of the human body to technology is an especial aesthetic theme (as admittedly it is with much contemporary work), we see no motivation for proliferating sensors and binding the performer's body to them. Certainly, in *Lightwork* we were trying to explore and suggest rather different images (both technically and aesthetically) of the relationship between technology and embodied gesture (a point we will return to at the end of this chapter).

Our discussion of V's experience in *Lightwork* also suggests that having *some* DM components, some directness to interaction, may be important for projecting certain kinds of gestures and making them legible to the audience. In particular, gestures which are punctate, associated with 'events', and which, say, announce the initiation or termination of some process may usefully be of this sort. Even here though, some critical gestural components will often need to be loosely coupled to any sensor system in use to allow for the unfettered expressive portrayal (exaggeration perhaps) of the DM components of the gesture.

Our reflections can be combined in the image of simple (i.e. avoiding sensor overload), loosely coupled (e.g. allow easy engagement and disengagement while enabling ongoing repairs), hybrid (algorithmically mediated and direct), gesture mediated (the identification of a gesture in some way influences how sensor data are further used) interaction techniques as one way of potentially reconciling aesthetic expressivity with technical effectiveness. What follows is a description of some of our deliberately primitive explorations under this rubric.

Gesture Mediated Interaction

We realise that our approach to gesture identification and processing runs against the grain of much contemporary work and may even seem trivial to some readers. For example, Modler and Ioannis (1997) discuss the use of time delayed neural networks to process multiple data streams from glove sensors and identify gestures therein, while Hofman and Hommel (1996) report on analysing data from an accelerometer equipped glove using discrete hidden Markov models. These are just two examples of the state of the art with a characteristic concern for specially designed, technically sophisticated and expensive peripherals yielding raw data subject to sophisticated mathematical machinery to identify gestures. The work on posture identification in Deliverable 2.1 is another example of this state of the art sophistication. Our approach is much cruder but—we believe—not just adequate to the task we have set ourselves (finding flexible, usable methods for gesture processing in experimental live artistic performance settings) but—arguably, of course—more appropriate given our emphasis (loose coupling and the rest).

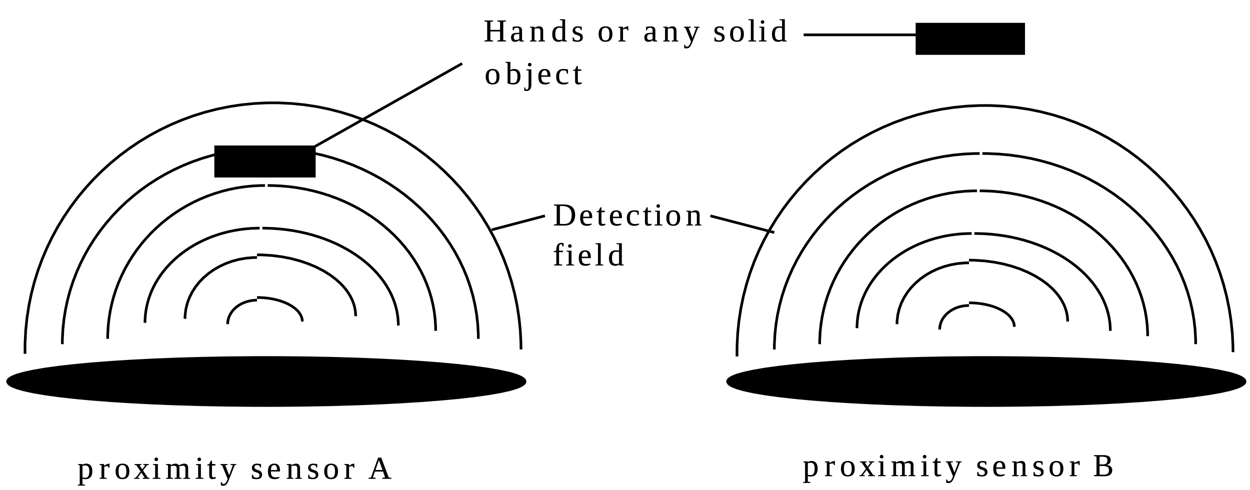
Our explorations to date have used the I-cube sensors and actuators marketed commercially by Infusion Systems and widely used by artists in performance and installation settings. The system consists of The Digitizer which transforms analogue electrical signals from peripheral sensors into MIDI data. The Digitizer can be programmed to a certain degree to configure, for example, sampling rates, sensitivities and the kind of MIDI data output (e.g. note data or MIDI controllers). Interface code exists to enable the Digitizer to be controlled remotely from the MAX programming language distributed by Opcode Systems or configurations can be downloaded to the Digitizer itself which can then operate in stand alone mode. A wide variety of sensors are made by Infusion Systems and support is given for users to construct their own. In what follows we describe our demonstrations with the I-cube proximity and pressure sensors and the simple datagloves they manufacture. The proximity sensors react to the presence of any object in a detection field and scale their output in relationship to the proximity of the object to the centre of the field. The pressure sensors respond to contact from a finger and indirectly estimate pressure by determining the area of the sensor in contact the finger. The datagloves have six similar pressure sensors within them, one for each finger tip, one mounted on the palm towards the base of the thumb. Clearly, the pressure sensors and the gloves require some contact to be made for sensor data to exist, while the proximity sensors are non-contact devices.

A Simple Example with Non-Contact Sensors

This example is one of our initial experiments in designing a sound controller for use in music performance (in particular for use by S in *Lightwork*). Our intention is to provide a very flexible gestural environment from very simple means. We use just two proximity sensors and give different interpretations (i.e. identify different gestures) to sensor data depending upon (i) how a data stream is initiated and evolves, (ii) which sensor is interacted with first, and (iii) whether or not the other is interacted with second in sequence. Many different gestures can be identified in this way and given qualitatively different interpretations in terms of sound control.

For example, below we describe a set up where each sensor can step through a predefined sequence of sound files every time you enter the detection field. The sensors can also control pitch, volume and filtering. The sensors are laterally placed (see Figure 2) with one intended for use by the left hand, one for the right. To clarify all this and to give a more detailed description let us say that sequence A associated with sensor A consists of sound files A0, A1, A2 and A3 and sequence B consists of sound files B0, B1 and B2.

In our currently preferred configuration, there are three operational modes.



F

igure: Two laterally placed proximity sensors.

1. If you enter the detection field of sensor A, it will then cause the sequence A to jump from sound file A0 (if A0 was the last sound file that was played) to A1 and start playing that sound file at a low volume. The volume will increase as you move closer to the centre of the sensor.

The playback of A1 can be further controlled by entering the detection field of sensor B. This lets you control the pitch of sound file A1. For example, the pitch may decrease as you get closer to the centre of sensor B.

2. If you first enter the detection field of sensor B, it will cause the sequence B to jump from for instance sound file B1 (if this was the last sound file that was played) to sound file B2 and start playing this sound file at a low volume. This volume will increase (as above) as you move closer to the centre of the sensor. To control the filtering you then enter the detection field of sensor A. This can be set up so that the filter cut-off frequency of a low-pass filter will decrease as you move closer to the centre of sensor A.

3. If you enter the detection field of either sensor A *or* B and *stay in the outer boundaries* of the field it will cause the same action as above, i.e. if you, for instance, enter the detection field of sensor A, it will cause

a jump to the next sound file in the sequence A and play that sound file at a low volume. If you now enter the detection field of the other sensor it will *also* cause its associated sequence to step forward and play the sound file in turn. Once you have entered the outer boundaries, and are thus playing two sound files at the same time, you can move closer to the centre of either or both of the sensors and control the volume of the associated sound file.

This means that in mode 3 you can play two sound files at the same time but only control the volume of these sound files whereas, in modes 1 and 2, you can only play one sound file but in return control the pitch, filtering and the volume.

Importantly, although one can identify a number of different 'modes' of operation using these two sensors, each mode is always available from a 'resting state'. It is not necessary to engage in further activity to 'switch' modes. The relevant mode is selected depending upon the *sequence and evolution* of activity the user-performer engages in. Furthermore, in this scheme, we do not have to await the 'parsing' or 'interpretation' of such a sequence before sound is heard. Sound is heard immediately on entering a detection field. It is the type of subsequent changes in the sound which are determined as a function of the mode that is identified.

Although we have found it convenient to speak of 'modes', our set up is *not* strongly 'moded' in the sense that this term is commonly used in the human computer interaction literature. The user-performer does not have to engage in an explicit action to achieve a mode or context switch which is external and additional to the action normally performed (cf. the use of modifier keys in conventional user interfaces). Different modes are identified but this is done through the interpretation of data streams within gestures, with the data streams themselves being operative throughout, selecting sounds, controlling level or filtering or whatever. In this way, we feel, a gesturally mediated approach to the manipulation of (here) sound can be technically effective yet flexible, and elegantly so. In particular, there is much scope for the performer to emphasise the gestural content of the performance over and above simply furnishing the detection fields with the presence of one or two capacitative objects.

Simple Whole Hand Gestures

This theme of giving an interpretation of sensor data in a manner which is sensitive to the context in which the data occurs, and doing this in a computationally simple way, has guided our work with the pressure sensitive gloves. Here, the 'context' is not so much a matter of identifying a sequence of activity (e.g. first A then B) but of identifying co-occurring simultaneous elements and adapting the processing of each accordingly. Again, more details of our exact example should make the point clearer.

After some simple thresholding to cope with noise from the sensors within the glove, different gestures are identified depending upon the sensor data streams which are currently active.

1. If no data streams can be detected, we identify a REST gesture.

2. If a stream of data can be detected coming from just one of the fingers (including the thumb), we identify what we call a POKE gesture. Clearly, there are up to five different instances of POKEs per hand which can be identified, one for each finger.

3. If a stream of data can be detected coming from four fingers (thumb excluded), we identify what we call a PUSH gesture. There is only one instances of PUSH per hand available.

4. If a stream of data can be detected from all five fingers and from the palm sensor, we identify what we call a PUNCH gesture. Similarly, there is only one instances of PUNCH per hand. (Note: we leave out the thumb from our definition of the PUSH to maximise its discriminability from the PUNCH).

5. If a stream of data can be detected from the thumb and also one from just one of the other fingers, we identify what we call a SIGN gesture. We envisage the user-performer opposing the thumb with one of the other fingers in this gesture. There are four instances of SIGNs per hand available for identification.

6. All other co-occurrences of data streams are classified as OTHER.

In our demonstrations so far, we have also placed one or two pressure pads (one per glove) on a rigid surface in front of the user-performer and used this to distinguish two different classes of POKEs (whether they are touching the pressure pad, POKET, or some 'background' surface, POKEB), PUSHes (PUSHT and PUSHB) and FISTs (FISTT and FISTB).

It is to be noted, of course, that there is not a one to one mapping between an actual physical gesture and the classifications we make. For example, if a user-performer presses the thumb and forefinger down simultaneously onto a rigid surface, this will be deemed to be a SIGN, as will the simultaneous pressing of those two fingers against each other in the manner hinted at by the name 'SIGN'. While one could argue that this is clearly a failure for our (very) bottom up techniques for gesture recognition strictly defined, we are not so worried. We would envisage artistic uses where, for example, PUNCHes, POKEs and the rest would be used for performance purposes by a rehearsed performer. Such 'pseudo-SIGNs', if they were not part of the performed gesture repertoire, would be unlikely to be deployed.

All of these different gestures (POKEB, POKET, PUSHT, PUSHB, FISTT, FISTB, SIGN as well as REST and OTHER, for that matter) are available for distinct interpretations. In particular, the sensor data from the glove elements (transformed by The Digitizer into MIDI) can be further transformed depending on the gesture it is part of. The gesture forms the context for further data interpretation. Exactly what further interpretation occurs can be made idiomatic in the light of the hinted meaning of the names we have given, though really the choice is arbitrary.

For example, a POKE gesture might be used to give some fine control to just one parameter in whatever interactive system is employed in the performance in question, e.g. the frequency of the cut-off of a low-pass filter. A PUSH suggests, perhaps, the manipulation of some 'bigger' mass of elements, say the relative levels of a number of simultaneous sound sources (perhaps four, one for each finger involved). A PUNCH suggests the most dramatic and, perhaps, violent of performance gestures. Here six streams of data become available for complex effects (perhaps governing the unfolding of an 'explosion' algorithm which multiplies sound sources in response to the overall intensity of the 'grip' of the punch or the magnitude of its impact on the further pressure sensor in the case of a PUNCHT). Finally, a SIGN could be suggestively used to indicate a transition from one set of sound sources and algorithms to another as a performance unfolded. Remember: once the gesture has been identified, all the available sources of continuous data within the gesture become deployable for real-time, moment-by-moment control.

As in the proximity sensor example before, an application analysing MIDI data from The Digitizer (which in turn has been configured for working with the appropriate sensors) has been authored in Opcode System's MAX programming language and forms the basis of our current demonstrations and early experience.

Provisional Conclusions and Future Work

In this section, we will review the status of work to date, document some problems encountered and suggest some future lines of research.

Overall, we feel our approach is promising, though our technologies are not mature enough yet to permit sensible formal evaluation. This is preliminary work which we shall deepen in Year 2 of eRENA. The methods we have implemented using the two proximity sensors work reliably and give intelligible and predictable control. Informal demonstrations of our work have received promising feedback from musicians and allied researchers after they have gained hands on experience with the system. We have noted more accomplished performers quickly develop coordinated patterns of gesture across the two hands to enable them to easily capitalise on the ways in which the system depends upon sequences of gestures. Much depends upon the coherence of the underlying sound file when it is being manipulated. If the sound file itself has discontinuities of some sort in it, this can make for a confusing impression if a user is making continuous, slow gestures over the sensors. In the hands of a musician acquainted with the and expressivity can be demonstrated. Nevertheless, it must be very clearly admitted that we are yet to undertake formal examination of these claims and that our claims must be treated with customary scepticism at this stage—especially as two of the musicians we speak of are ourselves!

In the case of the two proximity sensor system, our main problems have been peripheral ones due to the electrical functioning of the sensors themselves. These problems have been more severe with the pressure gloves which have rarely worked reliably, it being not uncommon to 'lose' a finger or two. This, of course, has disastrous effects for a gesture processing technique which requires a fully working glove for some of the gestures to be recognised at all. We understand that Infusion Systems are revising their glove designs. These matters aside, there are some more difficult issues with our approach to gesture processing in the glove case which do not appear when just two simple sensors are being considered.

Data from the gloves can be 'spikey'. In the transitions between gestures, for example, even the most careful user can make strong contact unintendedly with one of the pressure pads within the glove if the glove slips over the hand or gets crumpled as the hand closes. Simple thresholding is not enough to eliminate this problem without losing much meaningful data. In our interpretative scheme above, most such transient spikes do not cause any false positives of gesture classification. Each glove has six pressure pads which, at any one time, either are or are not active (i.e. are yielding data above threshold level). This yields two to the power six (=64) possible combinations of activity. Only 11 of these are classified as target gestures. No activity is deemed to be REST and the category OTHER absorbs the remaining 52 combinations. Thus, most random combinations of transient data will be classified as OTHERs and, if no critical events occur consequent upon these, our approach is quite robust in the face of transient noise. The only exception to this is the occasional false alarm of a POKE (which requires just one sensor to be active) when the REST state is expected. However, if POKEs are interpreted so that some non-linear scaling takes place on sensor values (e.g. so that low amplitude transients have little effect), little further sophistication of data smoothing is *necessarily* required to minimise the effects of transient glove data (though we shall revisit this point shortly).

More troublesome is a 'flickering' between gesture classifications which can sometimes be observed at gesture onset and occasionally during a gesture intended to be sustained. For example, if the user-performer intends to establish a PUSH but first makes contact with the forefinger a POKE will be identified. If the little finger (pinkie) next makes contact, the classification will switch to OTHER where it will stay as the remaining fingers get established in the gesture until PUSH is finally identified. If the user-performer then attempts some expressive further movements within the PUSH but during this releases one finger to below threshold momentarily, then again the classification will flicker through OTHER.

This problem can be addressed in a number of ways. Some filtering can be applied to the glove data to smooth such transitions. This may reduce the responsivity of the gesture classification in that, say, the onset and release of a gesture may be delayed in their detection (as smoothing would be implemented by weighting in an influence from recent past values along with current ones). But, if it is at the cost of a reduced rate of false positives or unintended identifications, this might be worthwhile. Alternatively, the classification could be 'smoothed'. That is, a value for the strength of the evidence for a particular classification (perhaps given as some weighted sum of all the input sensor data values) could be computed alongside the classification itself and this would enter into a similar smoothing process. Again, a momentary drop from a PUSH to an OTHER might be smoothed over this way. Also again, the trade-off between accuracy and responsiveness would have to be assessed. This could be computationally more efficient as the smoothing calculation could be triggered only when a candidate gesture transition has been detected, rather than on each sampling of sensor data.

Another approach would be to introduce some impression of the 'neighbourhood' around each target gesture into the smoothing process. For example, in the intended transition from REST to PUSH, unless all four sensors activate at exactly the same time (i.e. within one sampling interval of each other), classification is bound to pass through certain kinds of OTHER. These could be ignored. Finally, some more top-down information about the transitions between gestures could be used based perhaps on the analysis of multiple empirical instances or through mainstream work on gesture processing and segmentation. We feel that it is somewhat 'overpowered' for our simple gloves and our requirements suggested above for supporting gestures in artistic performance (rather than demonstrating generic gesture processing techniques and technologies).

Besides, intellectually, we remain perversely attracted to *thoroughgoing* bottom up techniques. It is our belief that, *provided* (i) a sensible gesture repertoire is given appropriate a priori definition, (ii) noise and other un-wanted artefacts such as transients in sensor data can be adequately thresholded, filtered or otherwise managed, and (iii) the user-performer is adequately trained-rehearsed in the gesture repertoire and in the disciplining of their own body to achieve the selections from the repertoire they wish for, and (iv) the subsequent processing of gesture data (in the context switched fashion we have been arguing for) is intelligible (and hence can serve as a useful source of feedback for the user-performer), *then* the techniques we have been investigating may be adequate for our target domain: artistic performance where the right combination is to be made of expressivity and control. This may seem like a lot of provisos but it does not seem to be a longer list than others we have encountered when gesture processing techniques are reviewed for their applicability to artistic performance! And anyway one should be suspicious if a shorter list appeared alongside work as preliminary as ours. Further experience with our techniques will be sought in Year 2 of eRENA. We intend this to be a mix of more formally analysed demonstrations and user-experimentation alongside concerted work in developing interaction devices for use with *Lightwork* and other artistic endeavours.

While our work is empirical in the broad sense of being based on practical experiences and trials (becoming more formally empirical as and when this is appropriate), our motives are equally as much aesthetic and conceptual. We are concerned to develop technical possibilities which allow for flexible couplings between the human body and technical apparatus. We do not wish to penetrate the body with various apparati or bind it within a shell of sensors no matter how bearable or wearable. We neither wish for this as performers, nor do we see in some of the cyborg forms that are commonplace in artistic work and theory an image of how human-technology relations are, will be, or should be. And we have this view, not because we are nostalgic for some fantasy of human domination of technology or for some false image of an existence free from technology. Rather we wish to explore performance and other artistic spaces which are heterogeneous human-machine ecologies, where humans coexist with technology in a loosely, lightly coupled way, where engagement and disengagement are options, and there is no motivation to raise questions of human-machine relations as if answers *had to be* given in terms of where the 'power' resides. To be sure, under some circumstances, we might seek control. On other occasions, we are prepared to delegate to machines. On yet others, we might be prepared to discipline ourselves to make things work at all. This suggests that we should not assume that there can be just one interaction paradigm for artistic performance. Direct manipulation techniques have their place alongside more indirect methods. Sometimes we may wish to algorithmically multiply user-performer gesture, sometimes compress it, othertimes maintain a more familiar one-to-one gesture to effect ratio. The current chapter is just part of this picture, concerned as it is with simple gesture processing techniques which might have complex effects. Whether this approach is called for very much depends upon the specifics of the artistic application. In our work, we intend to combine it with the more indirect techniques explored so far in *Lightwork* in a heterogeneous interaction environment. At the end of Year 2 of eRENA, we will be able to more confidently evaluate our experience.

References

Hofmann, F. and Hommel, G. (1996). Analyzing human gestural motions using acceleration sensors. In Harling, P. and Edwards, A. (eds.) Progress in gestural interaction: Proceedings of Gesture Workshop '96. London: Springer.

Modler, P. and Ioannis, Z. (1997). Emotional aspects of gesture recognition by neural networks, using dedicated input devices. In Camurri, A. (ed.) KANSEI: The technology of