KUNGL TEKNISKA HÖGSKOLAN

CID-182 ● ISSN 1403-0721 ● Department of Numerical Analysis and Computer Science ● KTH

# Mixed Reality Group Interaction

eRENA ESPRIT Project 25379 Workpackage 7 Deliverable D7b.4

**Bernd Lintermann, Jeffrey Shaw, Heike Staff, Ian Taylor, Jason Morphett, John Bowers, Alex Butterworth**

CID, CENTRE FOR USER ORIENTED IT DESIGN

**Bernd Lintermann, Jeffrey Shaw, Heike Staff, Ian Taylor, Jason Morphett, John Bowers, Alex Butterworth**

Mixed Reality Group Interaction

**Report number:** CID-182, eRENA ESPRIT Project 25379 Workpackage 7 Deliverable D7b.4
**ISSN number:** ISSN 1403 - 0721 (print) 1403 - 073 X (Web/PDF)
**Publication date:** August 2000
**E-mail of author:** bowers@nada.kth.se

**Reports can be ordered from:**

CID, Centre for User Oriented IT Design
NADA, Deptartment of Numerical Analysis and Computer Science
KTH (Royal Institute of Technology)
SE- 100 44 Stockhom, Sweden
Telephone: + 46 (0)8 790 91 00
Fax: + 46 (0)8 790 90 99
E-mail: cid@nada.kth.se
URL: http://cid.nada.kth.se

# Deliverable 7b.4

# Mixed Reality Group Interaction


## ABSTRACT

The deliverable reports on the research and implementation of a working group interaction mixed reality prototype that relates to the general notion of interactive cinema. Building on former eRENA research work, task 7b.4 integrates the project's mixed reality demonstrators with Inhabited Television while deploying various interaction techniques and principles developed by the partners (EVE, MASSIVE, BFinder and MTK).

The deliverable documents the two week workshop and the one day public presentation where recordings from "Avatar Farm" (the third Inhabited Television demonstrator) were used to constitute this group interaction immersive cinematic environment. The workshop's results are evaluated both in terms of its technological and group interaction effectiveness.


| | |
|---|---|
| **Document ID** | D7b.4 |
| **Type** | Report and Video |
| **Status** | Final |
| **Version** | 3.0 |
| **Date** | 16 August 2000 |
| **Author(s)** | Bernd Lintermann, Jeffrey Shaw, Heike Staff (ZKM), Ian Taylor (Nott), Jason Morphett (BT), John Bowers (KTH), Alex Butterworth (Illuminations) |

## Content

# 1. Introduction

## 1.1 The basic approach

In WP 7b.4 we have undertaken to research and implement a working prototype that relates to the general notion of interactive cinema. The basic challenge relates to modalities of audience interaction with a modular cinematic structure in terms of narrative (dis)continuity and multiple points of view. In particular the Avatar Farm mixed reality environment created by Inhabited Television offered us the context of a spatially distributed narrative that an audience can experience and interact with.

Central to the task was the issue of group interaction within such a spatially distributed narrative. Previous work done in eRENA, including the studies of artist's installations that embody group interactivity, have shown that this is a complex challenge that demands characteristic solutions, both conceptual and technological, if it is to succeed. In this WP we set out to explore and evaluate one specific set of conditions that would allow the simultaneous interactivity of a larger number of persons in a mixed reality environment to be effected.

This WP integrates and extends various components of work done by five partners during the eRENA project, both in terms of technology and production methodology:

- The narrative film material was derived from the third Inhabited Television Avatar Farm demonstrator developed by the University of Nottingham and Illuminations (WP 7a.3). After two days of recordings of this real time interaction between avatars and participants, specific sequences that embodied sufficient temporal coherence were selected to be used within the framework of this workshop.

- Four recorded sequences were selected that were temporally parallel but where each was seen from the differing virtual points of view of the four participants in Avatar Farm. These were projected into four quadrants of the ZKM's EVE inflatable dome[1].

- The live Avatar Farm interactions were recorded as MASSIVE files[2] and applied using the MTK (Mapping Tool Kit), a comprehensive software development made by the ZKM in the context of eRENA[3]. MTK was also used to apply the incoming data from BFinder's video analysis of the audience's physical behaviour.

- BFinder, a software technology developed by BT and KTH, was used to analyse and map the incoming data from the video camera that was situated at the apex of the EVE dome looking down at the physical behaviour of the audience.[4]

---

[1] see eRENA D 6.3 Individual and Group Interaction, ed Hoch, M. et al, 1999
[2] see eRENA D 7a.3 Third Demonstration of Inhabited Television, ed Benford, S. et al., 2000
[3] see eRENA D 4.3/4.4 Production Tools for electronic Arenas: Event Management and Content Production, ed. Bowers, J et al., 1999
[4] see eRENA D 7a.1 Evaluation Out of this World: An Experiment in Inhabited Television, ed. Bowers, J et. al, 1999, D 5.3 Motion Control of Crowds, ed. Morphett, J. et al, 1999

- This demonstrator also builds on the specific understandings gained from earlier eRENA studies on group interaction with mixed reality environments in WP 6.3[5].

### 1.2 The concept of the workshop and presentation

On July 21, we invited two groups of people for the presentation of the workshop in Karlsruhe. On each occasion approx. twelve persons viewed and interacted with the selected Avatar Farm sequences. These were 5-7 minute long sequences that were edited from the original two hours of recorded material.

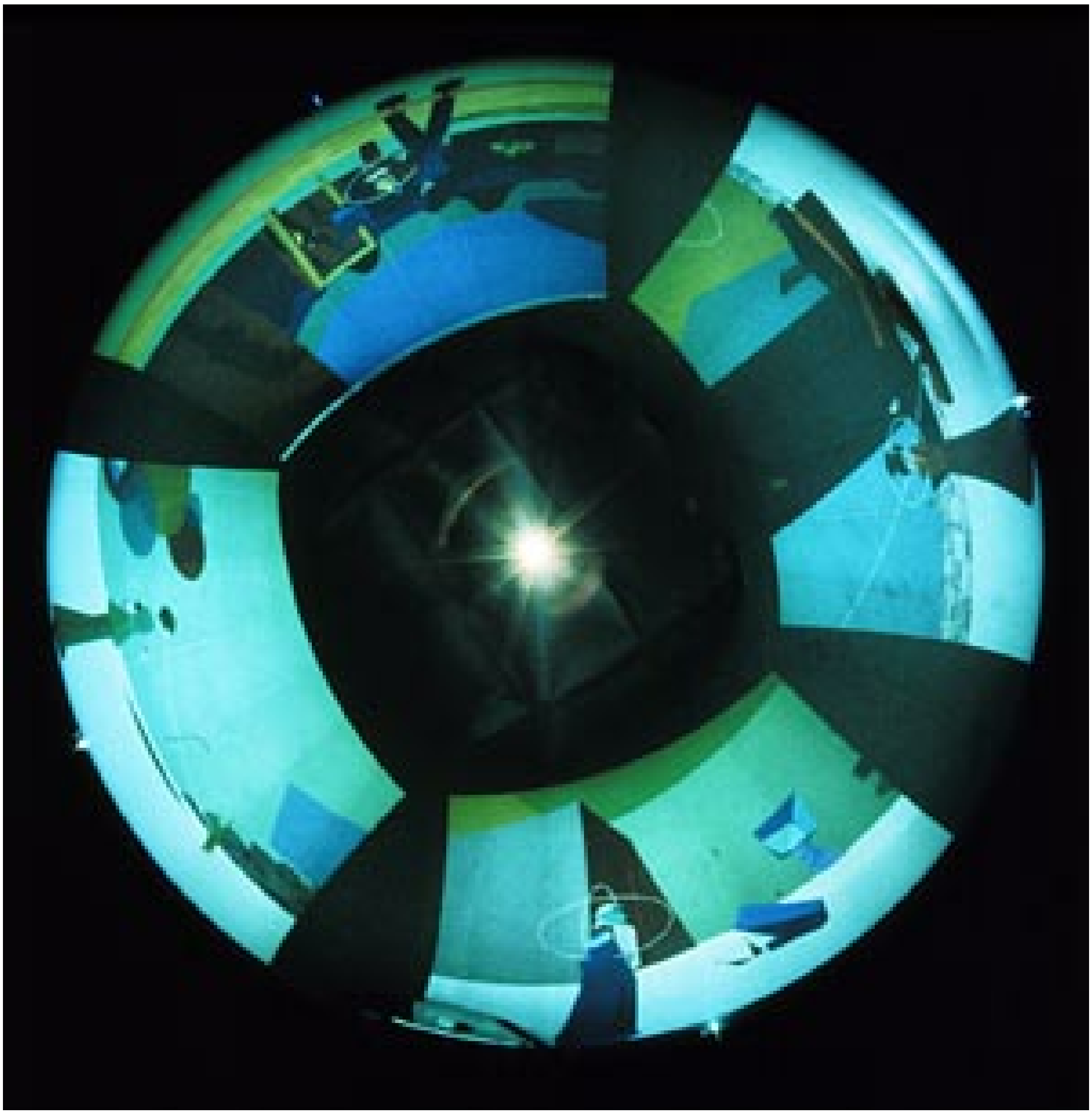


Figure 1

Four synchronically projected images in the four quadrants of the EVE dome showed parallel happenings in the Avatar Farm virtual world, each seen from the perspective of four different participants who were originally inside that virtual world.

5 see D 6.3 Individual and Group Interaction

By moving around inside the EVE dome environment (whose floor area was approx. 65 sq. m.) the visitors could influence the respective volumes of the four separate sound tracks emanating from the four projected sequences. In this way they could shift and focus their interest on one or other of the dialogues, which in turn would interactively modulate the overall narrative structure of these dialogues as well as the effective confluence of the Avatar Farm events.

# 2. Preparation of the Workshop

### 2.1. Working out of the first ideas

After the first broad concepts for this demonstrator were discussed in 1999 the five partners met in London on March 1st and 2nd. (Illuminations: John Wyver and Alex Butterworth, Nottingham: Steve Benford and Ian Taylor, KTH: John Bowers, BT: Jason Morphett, ZKM: Jeffrey Shaw, Adolf Mathias, and Heike Staff) and decided to adopt the following strategies:

- Narrative Content: four short stories to be edited out of the two days live recording of Avatar Farm
- Projection: 360°dome projection of the virtual world space showing four films running in parallel from four different participant points of view
- Group interaction: navigation and selection processes by tracking the spatial and temporal distribution of the people in the EVE dome
- Technology: should be based on the integration of eRENA technologies developed by the partners – BFinder, MTK, and MASSIVE

### 2.2 Avatar Farm

Avatar Farm is the title of an improvised interactive story that Alex Butterworth has written for the third Inhabited Television demonstrator[6]. The story is about a gods and demi-gods, who reside in a virtual world built out of four worlds - Kindergarten, Nirwana, Trade and Power, Behaviour Shift. The four participants find themselves thrown into this world of malicious and constantly intriguing gods, and in order to free themselves they have to learn how to manipulate the myths of their worlds.

On June 17 and 18 the live recording of Avatar Farm took place at Nottingham University. During these two days four chapters, each 30 minutes long, were performed with professional actors controlling the Avatars, as well as with four persons who had participated in earlier manifestations of Inhabited Television.

All the people involved in the planning and realisation of the Karlsruhe workshop were present at this recording session, either working with MASSIVE, directing the actors, or just as observers. On the day following these recordings, the definition of the workshop was finalised in accordance with the nature of this recorded material, and the first technical tests of the workshop procedures were begun in Nottingham.

The main features of the workshop that were here decided upon were:

- The audience would be able to control the respective sound levels of the four predominantly speech driven narratives in Avatar Farm. The focus on this acoustic interactive option was determined largely by the fact that MASSIVE was unable to do real time manipulations of the recorded visual material.
- The temporal consistency would be determined by the real time behaviour of the four participants in Avatar Farm. As most of the time they would be visitors in

[6] see D 7b.3 Third Demonstration of Inhabited Television

more than one of the Avatar Farm's four words, the four projections would present their differing locations and points of view. Only when they came together in one world would the representation switch to one shared panoramic point of view across all four projections.
- The physical distribution of groups of people in the dome should be the significant parameter of the tracking software, as well as the behaviour of individuals. For the purposes of evaluation at least two methodologies of interpretation and mapping should be implemented – one simple one (majority identification) and one with more complex parameters (both individual and group identification)

### 2.3 The workshop

The persons involved in the realisation of the workshop were: Ian Taylor (Nottingham), Alex Butterworth (Illuminations), Jason Morphett (BT), John Bowers (KTH), Bernd Lintermann, Torsten Belschner, Torsten Ziegler, Manfred Hauffen, Heike Staff (ZKM)

The on-site schedule was:
July, 10, 11. Physical installation of the various hardware components – CCD camera, infra-red lighting, video projectors, computers with various graphic and audio cards, Midi-interfaces, audio mixing desk, and the network.
July, 12, 13. Installation of the MASSIVE application and its integration with MTK and the projection system
July, 14, 15. Installation of BFinder and its integration with the CCD camera, MTK, and MASSIVE.
July, 16, 17. Functional integration of the sound equipment with the other components
July, 18, 19, 20. Final integration of the complete application, and tuning the interactive modalities of operation.
July 21: The public presentation

The workshop took place in the ZKM's inflatable EVE dome which was an environment ideally suited for the workshop's 360 degree projection and single overhead infra-red camera tracking system. Furthermore a coherent circular floor space for the visiting public to move around in was defined by this enclosure. From an architectural point of view this domed space also has specific experiential properties which support the dramatic articulation of a mixed reality environment such as was presented in this workshop.

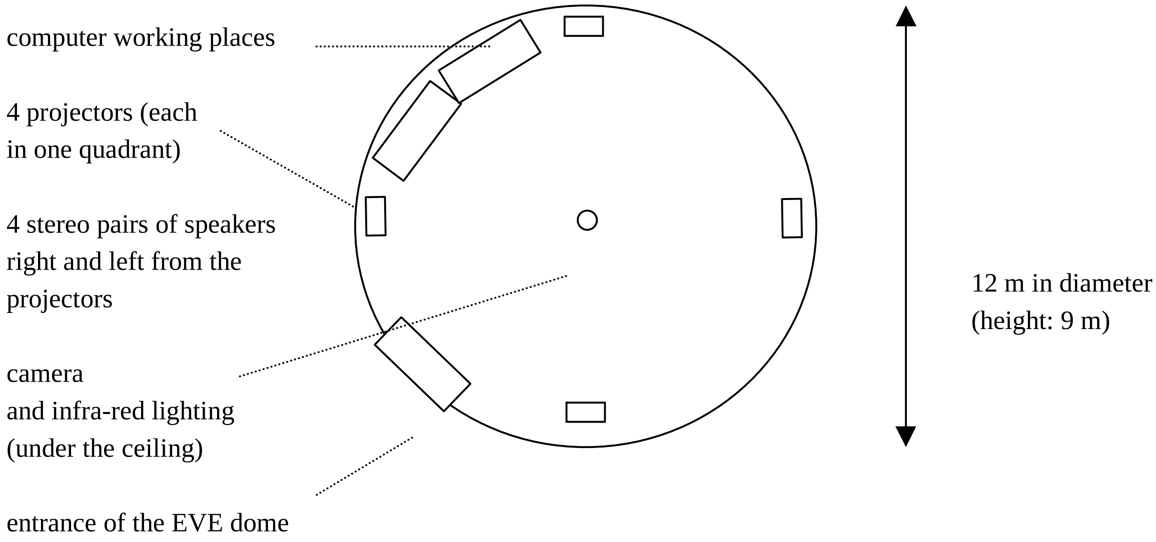



Figure 2: Ground plan of the dome

Figure 3: Workshop (from left to right: Alex Butterworth, Ian Taylor, Torsten Belschner, John Bowers)

# 3. Technical Set-up

### 3.1 General description

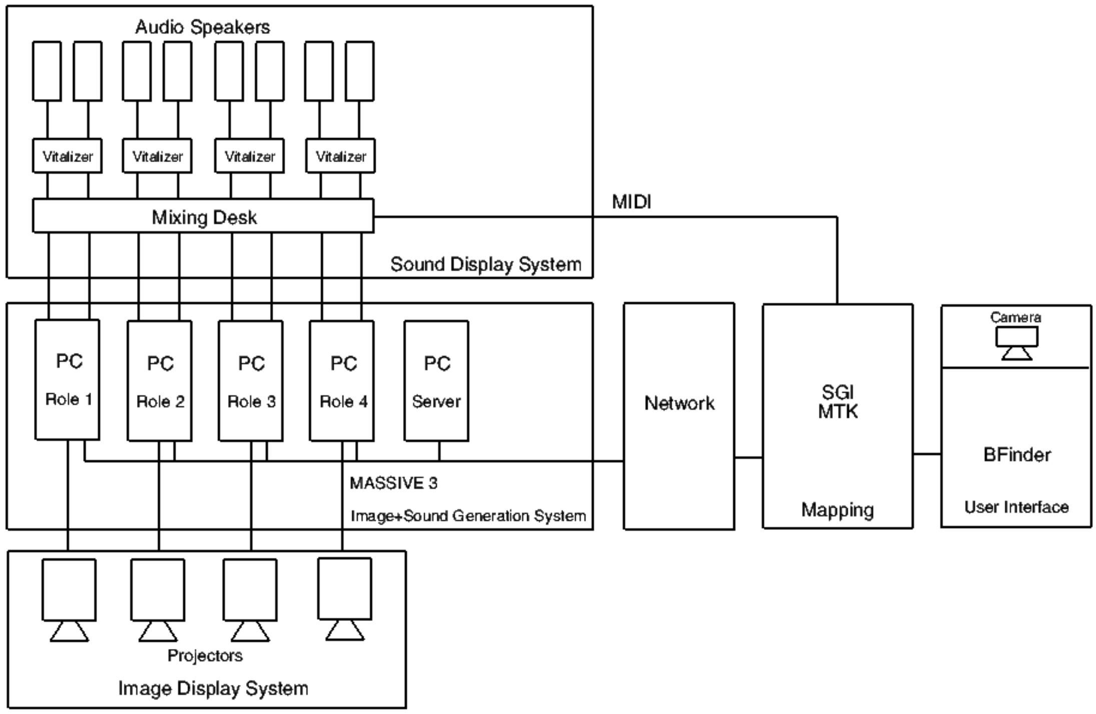



Figure 4: The components

The technical set-up in the EVE dome consisted of the following components:
a) Image display system: on the silver interior surface of the inflatable EVE dome using four Sony X1000 LCD projectors
b) Sound system: ACM mixing desk, SPL vitalizer, Genelec speakers
c) Image and Sound generation system: 5 PCs running MASSIVE 3
d) User Interface: 1 PC running BFinder, infrared CCD camera with compact infrared light source situated at the EVE dome apex
e) Mapping system: 1 SGI Indigo2 running MTK
f) Network: 12 channel HUB connecting the PCs and the SGI

In order to project the recorded scenery from the viewpoints of the four original Avatar Farm participants, the dome was divided into four quadrants, each quadrant showing the visual and audio apperception of one of these participants. The four video projectors were set up in a symmetrical layout projecting onto opposite sides of the dome. The eight sound speakers were located in pairs of two assigned to each projection area. The scenery was generated with the MASSIVE 3 software using 6 networked PCs. Four of these were assigned to Avatar Farm scene generation – one per projection. Two further PCs were used as servers for replaying the temporal links and distributing data.

The VGA video outputs of the PCs related to the scene generation were connected to the projectors, the stereo audio outputs were connected to a Niche ACM mixing desk, which is controllable via MIDI. The quality of the audio was enhanced using 4 SPL

MK2 Vitalizers. A CCD camera was mounted at the apex of the EVE dome, its wide angle lens covering almost the complete ground area. The camera was connected to a PC running BFinder, the propriety image processing software used to extract information concerning individual and group behaviour of people in the dome. This data was reported via TCP/IP to an SGI Indigo2 running the MTK software. In MTK the data was appropriately mapped to generate MIDI data for controlling the sound volume. A second MTK application running on the SGI invoking the creation of the participant viewpoints on the four PCs; and when all four viewpoints became proximate in the same virtual world. MTK generated appropriate views in order to achieve a full 360-degree panoramic image.

## 3.2 Technological components

### 3.2.1 Massive 3

The MASSIVE-3 set-up required 5 PCs to run successfully. Four of these PCs ran the client software, one for each role, whilst the fifth ran the processes required to generate the Temporal Link. ‘Temporal Link’ is the term used to describe the mechanism for replaying past recordings of virtual environments. When a recording takes place, a file is generated consisting of serialised C++ objects, which describes all geometry and interaction that takes place in the virtual world. This file may then be replayed at any later time to re-create the action exactly as it took place. If preferred, the replay could be subject to spatial and temporal distortions to achieve certain effects. For example, the replay could be viewed in slow motion, or in reverse, or the entire world may be viewed in miniature. In this case, no such transformations were applied and the action was to be viewed exactly as seen within the Avatar Farm event.

Five edited scenes had been prepared from the material that was originally recorded during the Avatar Farm. These scenes contained key events that took place throughout the development of the storyline, beginning with the Role’s first awakening in Nirvana and culminating in the imprisonment of the chief deity; Virbius. Viewed sequentially, these scenes provided a legible summary of the complete Avatar Farm recorded material.

When playing the recordings, a number of software components must be running. Figure 5 illustrates these components and the relationships between them.

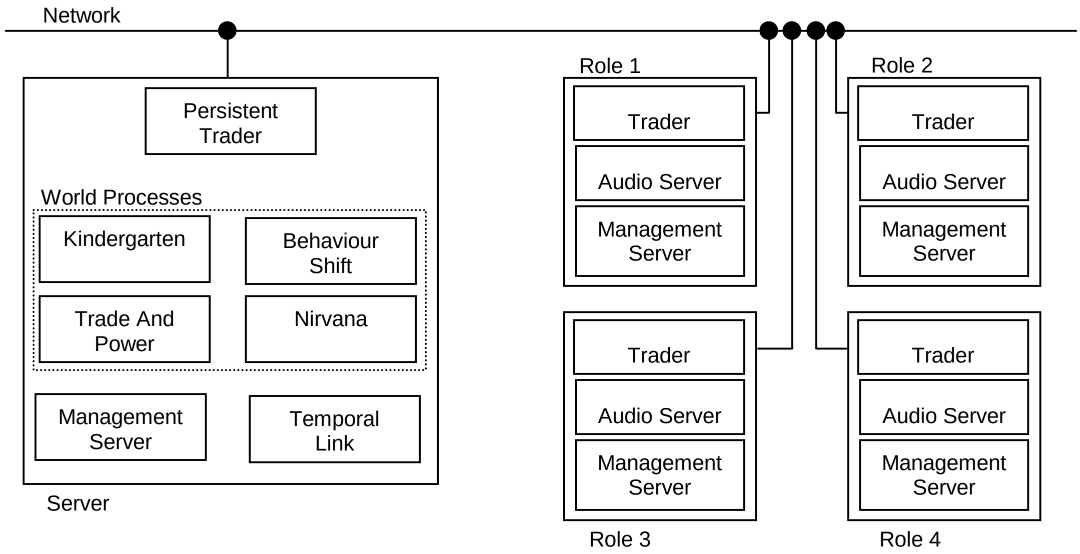



Figure 5: MASSIVE-3 processes

The function of the processes illustrated in figure 5 can be summarised as follows:

- **Trader/Persistent Trader.** The Trader is the standard naming service that is required to bootstrap any use of MASSIVE-3. The Persistent Trader is a variant on the Trader that supports persistent virtual worlds. It must be used on the Temporal Link replay machine as it enables restoration of a virtual environment from a serialised file.
- **World Processes**. When a Temporal Link history file is to be replayed, it must be given a destination in which to replay the content. Therefore, the Avatar Farm worlds are re-created to receive their original content. In this case, they are initially empty and are re-constructed when the history file is first read.
- **Management Server**. This is a MASSIVE-3 application that acts as a TCP server, allowing an external application to access a MASSIVE-3 world via a simple text protocol. The management server also supports rendering of 3D views under the control of the external application. It is through this application that both the Temporal Link Control Panel and MTK communicate with the worlds.
- **Temporal Link Control Panel**. This is a simple graphical user interface, written in JAVA™ that initiates the temporal link and offers simple playback controls (e.g. play, pause, re-wind).[7]
- **Audio Server.** In order to hear the audio on any machine, an audio server process must be started.

By using the Temporal Link Control Panel, the clips were first loaded and replay was started as each presentation began. Subsequent sessions involved deleting the content from the worlds (via the panel) and re-loading the content from the next edit files. During a presentation, each client machine rendered an over-the-shoulder view of the corresponding role. This view was instantiated by management server, which was acting under instructions received over the network from MTK. The integration of MTK and the MASSIVE Management Server is described in section 3.4.

[7] This is described in more detail in D 4.5 Production and Management of Events in Electronic Arenas.

### 3.2.2 BFinder

BFinder (Blob Finder) was developed specifically for the ZKM workshop and publicly demonstrated research carried out by BT on research into models for group analysis and identification carried out previously. BFinder was developed by BT (Jason Morphett) in collaboration with KTH (John Bowers). The origin of which was initially demonstrated in Work Package 5[8] in collaboration with another eRENA partner, EPFL. The interaction model was initially developed to manage scalable communications in large-scale virtual environments and used a generic algorithm based on the Spatial Model of Interaction[9] such that it could be applied in the vision system as well as earlier, in a 3D virtual environment.

The model's application to the BFinder vision system proved useful in identifying real-time group interaction in a non-invasive manner (~20Hz), reorienting the audio/visual projections in line with peoples grouping behaviour in the EVE dome. The system used the combination of an infra-red CCD (Charge Coupled Device) camera with the Dynamic Set (DS) model for group interaction. The DS model analyses entity positions in line with Hall's proxemic distances for social interaction[10] and creates a group entity – the Dynamic Set. This combination of features produced images such as that shown below in Figure 6 below.

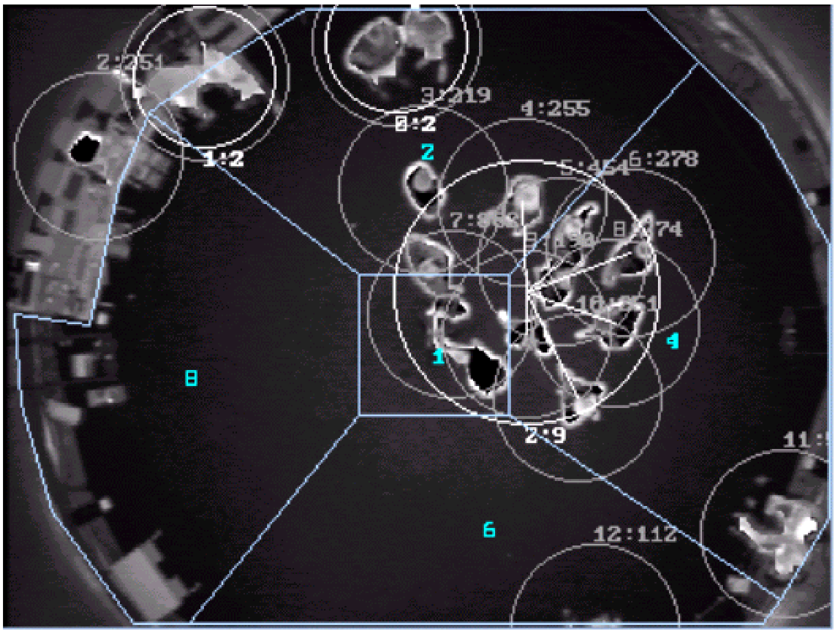



Figure 6: BFinder Main Window

Figure 6a is a top-down view as seen from the CCD camera onto the floor space in the dome. Each recognisable person (seen as an XOR'd 'blob') in the scene is identified through a personal 'aura' denoted by the grey circles around them. The label above each aura represents the blob identifier and its associated area in pixels (more on this

[8] eRENA D 5.3 "Motion Control of Crowds", ed. Morphett, J. et al, 1999

[9] Benford, S., D., Fahlen L. E., "A Spatial Model of Interaction in Large Scale Virtual Environments", Proc. Third European Conference On Computer Supported Co-operative Working (ECSCW '93), Milano, Italy, Kulwar Publishers, pp. 109-124

[10] Hall, E., T., "The Silent Language", 1963

pixel area later). A Dynamic Set is denoted through a predominate white aura as seen towards the middle of the scene. This aura has a label at the bottom showing its identifier and the number of *members* contained in the DS.

The cyan lines on the image represent the *regions*. These were used to partition the interaction according to the spatial distribution of DS's and individuals. For example, the normalised distribution of individuals in Figure 6 above shows that most are located in regions 2 and 4. The normalised distribution of DS's shows that most are in region 2, but the DS in region 4 is larger. This information along with the number of members in each DS and the total number of 'blobs' in the scene was sent over the network to MTK for interpretation and media routing as shown in Figure 4.

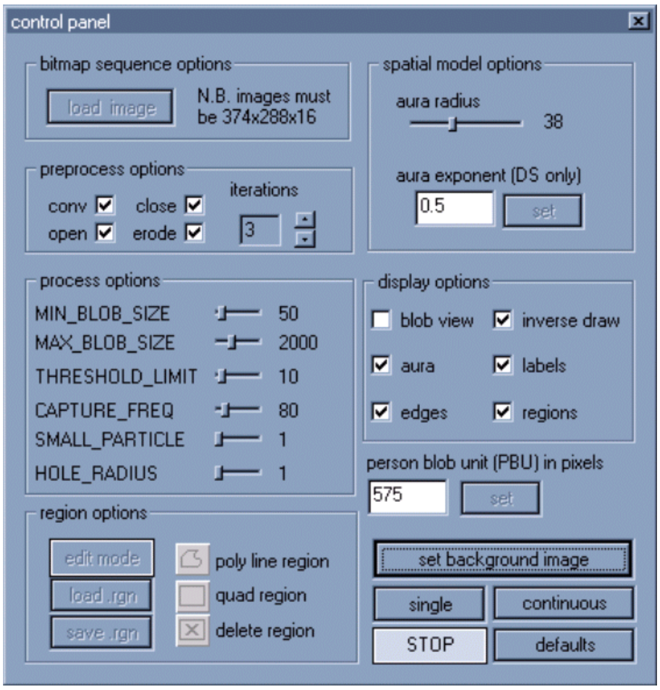


Figure 7: BFinder Control Panel

BFinder is controlled by a user interface panel (see Figure 7 above). This panel allows for the alteration of image processing and DS parameters in real-time. This real-time calibration allowed us to set-up and calibrate the software in a very short time frame. The results of the calibration are serialised to the registry such that restarting the application begins with a read of these values into the application during the boot-strap sequence the next time.

Factors such as minimum and maximum blob size as well as 'noisy' particle areas could be adjusted to clean the image before DS analysis took place. Another feature of the control panel was to allow the definition of the *PBU* (Person Blob Unit) to be defined at run-time. The PBU was used to counter the problem found when blobs touched each other. Because of the way the image processing algorithm for blob analysis works, touching blobs were 'closed' and treated as a single entity. In the workshop however, this was a problem. Therefore, by defining the PBU to be approximately equal to 1 ½ persons, we could inspect each blob and interpret it as being either a single or 'group' entity by dividing its blob area by the PBU. The effect of which can be seen in Figure 6 for DS's 0 and 1.

BFinder allowed other parameters to be defined at run-time including the point-and-click editing of the regions and their subsequent serialisation. This real-time calibration of the tool affords BFinder's employment in other domes or environments where non-invasive group based region tracking is required.

#### 3.2.3 MTK

In order to implement the functionality required for the workshop, MTK has been extended in several layers. First a generic concept has been developed and implemented for reading and writing data from and to arbitrary hardware devices. This concept abstracts from a concrete hardware device and offers a generic interface to MTK that means a set of operations where just the low-level operations have to be implemented for a specific device. It offers a generic database for storing data received or to be send, offers automatic parallel processing for slow hardware devices like the serial port in order to not slowing down the main application, it implements an abstraction layer for ASCII and binary data and introduces a protocol layer for higher level devices.

On top of this implementation a MIDI device node for the sound control and a TCP/UDP device node which supports network communication has been implemented. The MIDI device node reads and writes the full set of MIDI events and makes them available as MTK parameters. The TCP/UDP device node was used to implement the nodes, which communicate with the BFinder application as well as MASSIVE 3. The MASSIVE node sends interactively definable commands to a MASSIVE management server and stores a database of entities, which are reported by the management server. Since the reporting mechanism implemented in MASSIVE is command based, that means each request to MASSIVE is send as a command line of ASCII characters with the command name and parameter, as well the received messages consist out of keywords and parameter, MTK nodes have been implemented which allow to compose and decompose ASCII strings in a print and scarf style known from the C language, involving MTK parameters like values, strings, vectors and matrices. Using the MASSIVE node for communication, an MTK network has been created during the workshop, which first creates and then controls the views visible in the four projections. Four MASSIVE nodes keep track of the positions of the roles on the four PCs. The MTK application then computes the average positions and detects if all roles are in an adjustable distance. In case commands are composed and send to each MASSIVE server invoking views which result a full panorama. In case the roles disperse commands are composed to relink the camera to the roles positions and orientations.

During the workshop another MTK network has been created which allows to interactively change the weighting of the incoming data from BFinder in several ways such as allowing mixes of single person and group detection, enhancing the speaker volume which is in peoples focus and changing the system reaction after periods of low activity. A filter node has been implemented to smooth data delivered by a noisy source. This has been used to smooth the transition in the sound volume people's focus change to different quadrants in the dome.

# 4. The public presentation

## 4.1 The presentation itself

Figure 8: Photo from the presentation

Two groups (each of 12 persons) were invited to participate in this public presentation. Invitations were sent out to all persons working at the ZKM and to Karlsruhe University departments and companies working in the field of collaborative virtual environments (CVEs). Most of the visitors that came had already had personal experiences with the use of interactive installations. Half of them were programmers and some were experts in VR and CVEs. Most of the first group of twelve persons were more generally interested people; the professionals predominantly constituted the second group. The groups each spent about half an hour in the dome.

Both groups were introduced to the presentation and provided with the following information/instructions:

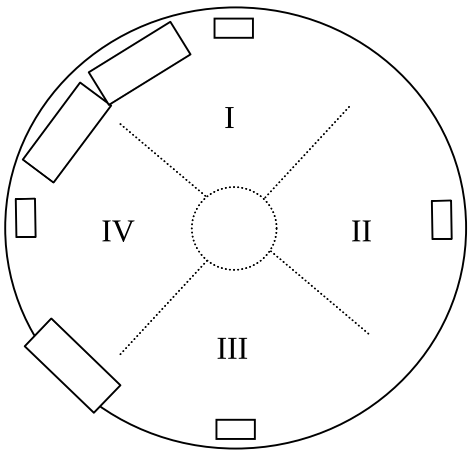
- They were told that Avatar Farm was originally an improvised play following a script and played in a CVE. Four 30 minutes chapters were recorded from which 5-7 minute long sequences had been selected for this presentation.

- They would see the four sequences consecutively in four separate projected scenes, each embodying the point of view of one of the original Avatar Farm participants.
- They were asked to explore how they could interact with this cinematic installation by moving around physically in the space.
- They were also informed that this installation was an eRENA research project, and that we were interested in all taking account of their experiences and observations for evaluation purposes.

Figure 9: John Bowers giving the introduction

**4.2 The tracking and interaction strategies used in the presentation**

Two different tracking and interaction strategies were applied in this presentation which we will call the A and B strategies. A was applied to the first and third Avatar Farm sequence, while B was applied to the second and fourth sequence. This arrangement did not relate in any way to the content of these sequences, but was simply a way of alternating the two tracking/interactivity modalities.



In the A strategy, BFinder and MTK were looking at all the individuals in the dome, and determining the relative density of persons in each quadrant. If the distribution was equal (say 6 people in quadrant 1 and 6 people in quadrant 3) then the sound levels would be symmetrically distributed (i.e. in this case 50% volume in these two quadrants). If the distribution of people in the various quadrants was uneven, then there would be a non-linear distribution of sound volume levels that emphasised the quadrant(s) with the larger number of people, and reduced the importance of the quadrant(s) with lesser numbers.

In the B strategy, BFinder and MTK were looking at groups of people as the significant unit. Two or more people who were closer than 1.6m to each other constituted a group. The tracking software analysed and gave relative values to both the size and quantity of such groups in the four quadrants, enabling MTK to accordingly modulate the audio levels associated with the four Avatar Farm sequences. In this B strategy another controlling feature was introduced: in the case that all these groups would become static for a period of time, the tracking software would shift its attention to any individual that left a group and moved into another quadrant, and this caused a temporary (approx. one minute long) audio level response to his new position. If a group did not consequently gather around this individual, the system would revert to the group-tracking mode.

### 4.3 Behaviour of the visitors during the presentation

The two groups clearly behaved in different ways. The first group of generally interested persons looked around, but started moving around only after some minutes. The second more experienced group of visitors immediately moved about in order to find out how to interact with the installation. The first group could not decipher the modalities of interaction until persons from eRENA mingled with them and gave them some hints by moving in a co-ordinated manner between the quadrants. The second group recognised how to interact during those sequences using the A strategy of tracking and interaction, but were clearly puzzled when the B strategy was applied.

A probable impediment to the visitors having an intuitive understanding of our organisation of the floor area into four quadrants (that were respectively linked to the four separate projections), was the circular nature of the dome which did not offer a clearly perceptible definition of these interaction zones. While none of the visitors remarked on this, this was a judgement we made when observing how people behaved inside the dome – they tended to circle around the circumference rather than fully exploit the functionality of the quadrant areas.

In general one could characterise the general behaviour of the visitors by saying that while they expressed an active and often successful curiosity about the group interaction modalities that were being offered, the situation did not provoke them to fully decipher and engage in an exact articulation of these interaction modalities. Instead these visitors tended to wander around and simply enjoy the chance effects of their uncoordinated group behaviour, which in itself generated a pleasing enough confluence of audio-visual results.

Figure 10: Photo from the presentation

### 4.4 Reactions of visitors after the presentation

Both groups were interested to discuss their experiences and our objectives. These discussions each lasted about half an hour. The reactions, questions and critical remarks are structured here under three headings: their understanding of the installation and the Avatar Farm material, the group interaction concept, and proposals for improvement.

The overall concept of the installation became quite clear to the majority of visitors. They also understood why it is necessary to make a choice when you have four parallel happenings and one wishes to focus on one set of dialogues. But due to the 8 bit audio-sampling rate which MASSIVE uses (compounded by the idiosyncratic acoustic properties of the EVE dome) the sound quality was not good, and many people had difficulties following the dialogues. Furthermore these were spoken in English which for most visitors was not their native language. These factors made it difficult to understand the content of Avatar Farm in terms of precisely identifying the roles and relationships between the avatars.

One visitor said that he was bored by CVEs where persons usually have simplistic exchanges via their avatars, but in the case of Avatar Farm he found it was great improvement to have a CVE that was structured by a prepared script within which the improvised interactions could take place. Most visitors very much liked the design quality of Avatar Farm, but at least two persons couldn’t decipher the reason for the white circles around the heads of the avatars (which indicate who is speaking or attempting to speak).

There were two persons, one from each group, who posed several questions concerning the group interaction concept and expressed critical remarks. Both doubted the basic notion of a group interaction where persons have to jointly co-ordinate their movements. While in a single user setting one has direct control over the parameters of interactivity and can individually unravel and engage the ‘rules of the game’, being in a group means one is dependant on the fortuitous circumstances of other peoples behaviour, making it very difficult just to find out how everything works. They also pointed out that such a group interaction approach was very much hindered by the social discomfort of having to co-ordinate with a group of strangers – maybe it would be easier if they were with a small group of friends. One said: "I must be very attracted by the whole thing in order to integrate myself with a bunch of unknown persons." (In the opinion of the eRENA partners, this apparent sensitivity to social engagement is a topic that may have also have a psycho-geographical dimension – one might get quite different results in other European regions.)

Most recommendations for improvement followed from the group interaction problems indicated above. These were typically: to reduce the parameters of interactivity to a bare simplicity, to give more extended information and/or hints to the visitors about how to control the environment, and one person proposed that there should be a guide that actually co-ordinates the behaviour of the visitors. The need to improve the overall audio comprehensibility was self-evident.

Figure 11: Discussion after the presentation

### 4.5 Conclusions

Entering the EVE dome environment and seeing four large immersive projections of a well designed mixed reality environment made a strong impression on the audience, and certainly invited them to actively engage in and explore its interactive narrative possibilities. The concept of multiple points of view within four quadrants of a 360-degree projection environment also proved to be an effective way of presenting this mixed reality environment, quite distinct from the usual desktop embodiments of CVEs.

Unfortunately the poor sound quality made it difficult for our visitors to follow the dialogues which from a dramatic viewpoint were a central feature of the Avatar Farm material that was used. This is especially the case considering the limited visual and gestural expressiveness that can be embodied by computer generated avatar figures, no matter how well designed they are. Furthermore the inherent excitement of a real time CVE (that was present during the recording sessions) is to a large extent lost when this material is presented in a linear playback mode. The improvisational dialectic becomes a documented performance, and the spectators tend to focus all their attention on the reading and hearing of this material so as to decode and reconstruct the meaning of the original event. This seems to indicate that in a group interaction mixed reality environment it might be preferable to constitute an audio-visual event whose narrative identity is actually constructed on the spot by the actions of the group.

Even for an experienced visitor it was difficult to simultaneously decode the content and explore the interaction parameters. We assume that some of these difficulties could have been avoided by offering them in advance much more detailed information about the nature of the installation and its group interaction functionality. Our hope that a group interaction environment could be as easily decoded as a single user environment proved to be unfounded – clearly the discomfort involved in initiating social exchange between the visitors worked against the co-ordinated effort that was needed to discover these interaction parameters. The quality of the material that was presented, the comparative simplicity of the interaction parameters, and the relatively small number of persons in the group did not in itself overcome this problem.

Group interaction has been one of the central eRENA research themes. It was explored in WP6[11] using non-narrative computer generated materials - flocks of fishes and four cloned human beings - and the interaction was done with small groups of people. These were simpler setting than the one we explored in this workshop. But probably the most important feature that contributed to the success of these earlier experiments was the direct reciprocity of activity in the virtual world with the activity of the persons in the real world. While quite simplistic from a narrative point of view, this correlated interactivity clearly allowed a number of people to achieve a sense of satisfying coherence in their relationship to each other and to the virtual configurations. This same feature has also been remarked upon in the studies of multi-user art installations, of which Perry Hoberman's Bar Code Hotel[12] is a paradigmatic example. In comparison, the workshop revealed the problems of introducing an autonomous set of narrative elements into the circumstance of group interaction, because as soon as the immediacy of directly correlated events between the virtual and real world was less evident, the viewers found themselves in a less transparent and more difficult to manage situation. Clearly there is a challenge here for future research in this field – how to conjoin these two components: pre-scripted narrative elaboration and the clarity of directly correlated interactivity.

[11] see D 6.4 Individual and Group Interaction

[12] see escape Deliverable 1.1 "Developing a framework for e-scapes" in: Presence and Representation in Multi-Media Art and Electronic Landscapes, ed Buscher/O'Brien/Hughes/Mariani/Rodden/Trevor, Lancaster 1998, p.113ff

Lastly there is the issue of group size and the mapping of group activity. Undoubtedly smaller numbers of people are easier to design for – in some ways one can consider a small group just in terms of a collection of individuals and still make use of single user strategies. With larger groups one is forced to subsume the individual to some extent, or even completely. The most obvious strategy of letting the majority decide might attract some people but puts off others – there are psychological, social and ideological issues involved here. But this simple strategy has a transparency that does allow group interaction to work – even with vast numbers of participants. On the other hand if one wants to affirm the individual as a more significant unit at the expense of the crowd, the level of complexity that is engendered when that crowd sees it has to break up and become individuals to have an effect on the work creates a mapping problem that can seriously debilitate any attempt at interaction transparency. Our workshop has certainly further elucidated these fundamental problems, but it also indicates that there is much scientific value and public enjoyment to be found just in the process of searching for tentative solutions.

## 5. Summary

This WP sets out to address two objectives set by the eRENA. Firstly that of the notion of group interaction with mixed reality environments, and secondly the creation of innovative forms of mixed reality environments – in this case an interactive immersive cinema.

Because it is a final year demonstrator we also set ourselves the task to make this workshop an integrative undertaking that would draw on and further extend the technologies and experiences developed by a majority of the partners during the eRENA project. In some respects this also limited the scope of what this workshop would achieve with respect to its group interaction objectives, because we chose to work with materials that had been initially developed in other eRENA contexts, rather than creating completely new materials specifically for this task. But we felt the integrative aspect was of greater importance at this stage in the eRENA project, enabling us to conjoin and make certain conclusions about the overall scope of work we had done till now. Understood in this way, this workshop is a convincing demonstration of the robust success of specific technologies developed in eRENA (BFinder, MTK, Inhabited Television, EVE, MASSIVE), not just as standalone tools and applications but also as components which can be interconnected and interrelated to work together to constitute a mixed reality application.

By inheriting and extending the datasets created by Inhabited Television in WP 7a.3, this workshop showed how that material could be scaled and adapted to a new form of mixed reality environment – moving from a networked desktop/television context into a large scale interactive and immersive cinematic experience.

The specific group interaction modalities that we implemented and evaluated are certainly not conclusive statements on this broad, challenging and precarious topic. Rather they constitute a research exemplar that furthers our understanding of the issues involved and of possible strategies of implementation. The effectiveness of the specific technologies we developed to track, map, and apply data with respect to the movement and distribution peoples physical presence in a mixed reality environment are certainly viable constituents of future work in this field, where the major challenge remains the articulation of audio-visual content structures that can suitably and elaborately respond to this data.