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

CID-83, KTH, Stockholm, Sweden February 1999

Evaluating Out Of This World: An Experiment in Inhabited Television

Steve Benford, John Bowers, Mike Craven, Chris Greenhalgh
Jason Morphett, Tim Regan, Graham
Walker, John Wyver

CID
Centre for
User Oriented IT Design

**Author: Steve Benford, John Bowers, Mike Craven,
Chris Greenhalgh, Jason Morphett, Tim Regan
GrahamWalker, John Wyver**

Title: Extended Performances: Evaluation and Comparison Report number: CID-83
ISSN number: ISSN 1403-0721
Publication date: February 1999
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/

Deliverable 7a.1
Evaluating *Out Of This World:* An Experiment in Inhabited Television

---

ABSTRACT

This document forms Deliverable D7a.1 of the eRENA project of the i3 schema of the ESPRIT-IV research action of the European Communities. eRENA is concerned with the development of electronic arenas for culture, art, performance and entertainment in which the general citizen of the European Community might actively participate supported by advanced Information Technology (IT). Within this general context, the current deliverable reports on the project's first experiment in 'Inhabited Television' whereby Virtual Reality (VR) technology is deployed to enable the citizen to participate within broadcast TV.

The document describes how, through an actual public demonstration, we were able to realise a popular format TV show within a multi-user virtual environment by paying special attention to the design of simple yet powerful interfaces for participants and viewers, together with novel camera control and event management software support for the production team. We have also subjected our work to extensive evaluation from a number of perspectives, both social scientific and technical. The results of these evaluations and their implications for work in the project and for more general research topics are described.

In particular, it is emphasised how this deliverable is the product of international, cross-partner and cross-profession collaborations in the project and how the work within it has been able to lever new integration pathways within the project.

Document	D7a.1
Type	Deliverable
Status	Final
Version	1.0
Date	28 February 1999
Author(s)	Steve Benford, John Bowers, Mike Craven, Chris Greenhalgh, Jason Morphett, Tim Regan, Graham Walker, John Wyver (BT, KTH, Illuminations, Nottingham), edited by John Bowers

Deliverable 7.1a:

Evaluating *Out Of This World:* An Experiment in Inhabited Television

Contents

Deliverable 7.1a:

Evaluating *Out Of This World:* An Experiment in Inhabited Television

Preface

John Bowers

Royal Institute of Technology (KTH), Stockholm, Sweden

0.1 Document Overview

This document forms Deliverable D7a.1 of the eRENA project of the i3 schema of the ESPRIT-IV research action of the European Communities. eRENA is concerned with the development of electronic arenas for culture, art, performance and entertainment in which the general citizen of the European Community might actively participate supported by advanced Information Technology (IT). Within this general context, the current deliverable reports on the project's first experiment in 'Inhabited Television' whereby Virtual Reality (VR) technology is deployed to enable the citizen to participate within broadcast TV. The document describes how, through an actual public demonstration, we were able to realise a popular format TV show within a multi-user virtual environment by paying special attention to the design of simple yet powerful interfaces for participants and viewers, together with novel camera control and event management software support for the production team. We have also subjected our work to extensive evaluation from a number of perspectives, both social scientific and technical.

0.2 Inhabited Television

In Inhabited TV, a multi-user, collaborative virtual environment (CVE) is in some way integrated with broadcast technologies to accompany a TV show and, in the more interesting cases, to provide some of a show's content. In what perhaps are the paradigmatic proposals for Inhabited TV, the show itself is set live within a CVE—with views of the real-time activity of participants being broadcast and viewers at home interacting, perhaps via set-top boxes, with ongoing events. As we put it in Chapter 1 of this deliverable: "Inhabited TV combines CVEs and broadcast TV to create a new medium for entertainment and social communication. The defining feature of this medium is that an on-line audience can socially participate in a TV show that is staged within a shared virtual world. The producer defines a framework, but it is the audience interaction and participation that brings it to life". Proposals for interactive TV and claims to be exploring a "new medium" are themselves nothing new. It is the combination of VR and broadcast TV which gives Inhabited TV its technical specificity, and the concern to support the participation of and social interaction between citizens in what has traditionally been a less accessible medium which gives Inhabited TV its i3 interest.

In this deliverable, we describe a series of experiments with Inhabited TV, beginning with the *NOWninety6* poetry performance, *The Mirror* and *Heaven & Hell – Live*. These early experiments raised fundamental questions for Inhabited TV concerning the extent to which it is possible to establish fast-pace social interaction within a CVE and to which it is possible to produce a coherent and engaging broadcast of this action. A fourth more recent experiment, *Out Of This World (OOTW)*, directly addressed these questions and forms the main topic of this deliverable. We describe how the formulation of Inhabited TV design principles, combined with the use of dedicated production software, for scripting and directing a show and for controlling virtual cameras, enabled us to create a fast-moving and more coherent experience.

0.3 Evaluation and Reflection

In past work in eRENA on Inhabited TV, we have devoted much effort to self-critical reflection on our efforts and ensuring that demonstrators of the Inhabited TV concept are carefully designed so that we learn from prior experience as we accumulate personal expertise and scientific results. The current deliverable marks a further intensification of this effort. The core partner grouping responsible for Inhabited TV in eRENA Year One (University of Nottingham, Illuminations Television and British Telecom) has been added to with explicit social scientific expertise from KTH (The Royal Institute of Technology, Sweden), this partner conducting a dedicated ethnographic field study of the production of *OOTW*. The addition of KTH to the partnership also ensures interest in Inhabited TV from partners within eRENA drawn from more than one member state of the European Union. The KTH study has enabled the more formal documentation and analysis of audience reaction to *OOTW*, as well as the suggestion of new technological developments arising from the study of the production process—developments which will be further pursued at KTH and by other partners.

In addition to the evaluation input provided by this reconfiguration of relations between partners in eRENA, researchers at Nottingham and British Telecom have continued to critically reflect upon the usability of their own technologies. In this deliverable, where appropriate, this reflection has been complemented by quantitative statistical analysis of the demands users place upon supporting Inhabited TV systems and networking arrangements. This is enabling us to piece together a picture of what collective activity in an Inhabited TV CVE is like so as to, for example, begin to inform the dynamic allocation of networking resources. This importantly complements KTH's social scientific study as we now have information about Inhabited TV participation that can inform technical design decisions as well as how we might go about creating broadcasts that are satisfying experiences for those concerned.

On the basis of these evaluative reflections, we feel we are able to conclude that our experiments clearly demonstrate the technical feasibility of Inhabited TV. Though we have suggestions for improvements, the event management and camera control software worked well. KTH's ethnographic work also argues that Inhabited TV is practically viable in another sense: *OOTW* provided ample evidence that the different professions which need to be involved in Inhabited TV (VR researchers, TV personnel) can indeed work together. Inhabited TV is not such an outlandish proposal that it has no resonance with existing TV practices or presents collaborative challenges which people cannot humanly meet. Quite the contrary, TV personnel (directors, producers, technicians) could see how their skills could be applied and, where necessary, modified with tolerable ease in the Inhabited TV context.

While we feel we are able to strongly argue for the technical and practical success of *OOTW* and, through this for the technical and practical viability of Inhabited TV in general, it is clear that greater attention needs to be paid to developing appropriate formats and content for this new medium before it becomes truly engaging. Audience feedback revealed that *OOTW* was not a particularly entertaining show and our selection of a game-show format (though justifiable on other grounds) was found to be disappointing from an aesthetic point of view. However, we feel that audience reaction and our own reflections have given us many suggestions for how to rectify this. In particular, as well as involving script-writers at an earlier stage of production, we are seeking to capitalise upon the creative strengths of on-line communities themselves in developing content for future demonstrators. In this way, we wish in future demonstrators (to be reported in Deliverables D7a.2 and D7a.3) to explore strategies to give the general citizen an even deeper participation (the co-authorship of content) in broadcast media through Inhabited TV.

0.4 Structure of this Document

After this preface, the document opens (Chapter 1) with a review of our previous experiences with Inhabited TV, offering a comparative review and laying down a series of research challenges. Chapter 2 describes the *Out Of This World* (*OOTW*) show and enumerates the fundamental goals it was intended to meet. A series of design principles for *OOTW*, which subsequent development and production work followed, is given. Chapter 3 focuses on the event management and camera control software that was developed for *OOTW*, describes how these were used, evaluates their success, and makes some suggestions for future development. Chapter 4 describes the 'WobbleSpace' video analysis technology that was used in *OOTW* to support audience participation. WobbleSpace is evaluated both from a technical perspective and in terms of its success in facilitating audience involvement. Chapter 5 concerns a series of quantitative and qualitative analyses of network traffic and user activity in *OOTW*. The degree to which *OOTW* was able to encourage activity amongst participants is compared to previous applications of CVEs where the same analyses were carried out. Chapter 6 reports the ethnographic study made of production work in *OOTW*, documenting in some detail how the show was realised by the multi-profession team that put it together. Chapter 7 opens with an analysis of audience reaction to the *OOTW* shows against the fundamental goals *OOTW* set itself. This is followed by further discussion of the *OOTW* experience for evaluating the technologies involved in it (event management, camera control, audience interaction). The chapter (and this deliverable) closes with a sketch of the approach we will be taking in future demonstrators.

0.5 Relationship of this Deliverable to the eRENA Workplan

Formally, this deliverable arises from Task 7a.1 (Evaluation of *Out Of This World*) of Workpackage 7a (Inhabited Television) and its delivery constitutes the completion of this task. Workpackage 7 is devoted to the development of public demonstrators of technology and results of the eRENA project through events, exhibitions, installations, performances and broadcasts which are open to the general citizen of the European Community. Workpackage 7a specifically concerns Inhabited TV and is structured around three demonstrators of which *OOTW* is the first.

The delivery and evaluation of *OOTW* naturally follows from the early development and design work for it which was conducted in Year 1 of eRENA and previously reported in Deliverable D3.1. Several details of the design of the show were revised between May 1998 (when D3.1 was authored) and the show itself in September 1998, the finalised design being documented here. The current document also builds considerably on what we were able to say then about the event management, camera control and audience interaction software which were given only cursory description in D3.1. Thus, this deliverable goes some way beyond what would be suggested by its title. Not only is *OOTW* extensively evaluated from a variety of perspectives, we also take the opportunity to report here on technical development work conducted in the interim since D3.1.

Much of this development work draws on emerging results from other workpackages in the project. For example, the event management and camera control work is properly part of Workpackage 4 and the audience interaction technologies are drawn from Workpackage 6. We have taken the decision to accelerate the formal delivery of these results from those other workpackages into the current deliverable, and this for a number of reasons.

First, it makes this deliverable more coherent for the reader if a detailed description of our technologies is given alongside their evaluation. Second, it demonstrates the relationship between these workpackages in action. While Workpackage 7 is devoted to demonstration, what it demonstrates precisely is technologies from Workpackages 4, 5 and 6. We felt that this relationship was made clearer if we included descriptions of, in the case of this demonstrator, work from Workpackages 4 and 6 here. Third, progress has been very good and we feel that we are in a position to accelerate delivery of descriptions of some of our Workpackage 4 and 6 technologies without compromising what those workpackages will be able to deliver at the end of Year 2.

Finally, we have taken the opportunity of giving a detailed deliverable at month 18 to demonstrate to the reader the new collaborative relations that exist in the eRENA project between partners and how reflections on experience are not merely shaping future work on Inhabited TV alone but are impacting and drawing on other workpackages. In all these respects, we hope to show how the project has been responsive to the recommendations of the Year 1 review.

Exactly how this evaluation of *OOTW* is anticipated to impact on future work in the project is discussed in detail in Chapter 7. Let us give a summary here though. Naturally, the experience of *OOTW* will directly influence future Inhabited TV demonstrators of which there are two in Workpackage 7a. The simultaneous participation of two partners in both Workpackage 7a and 7b (Nottingham and KTH), both of which deliver at month 18, has enabled a number of cross-project issues to be identified. Chapter 7 (in 7.7) contains an explicit discussion of one of these and describes how this will receive explicit attention in future contributions to Workpackage 4. In this way, an integration pathway has been opened up between Workpackage 4 and *both* demonstrators.

Just as this deliverable reports on results drawing on Workpackages 4 and 6, so do its results set the agenda for future work in these workpackages. For example, Chapters 3 and 6 contain evaluation of the application for controlling virtual cameras which has emerged from Workpackage 4 and was used in *OOTW*. Chapter 7 gives an indication of future enhancements which could be made and how to better support the real-time direction of Inhabited TV based on this evaluation. Technologies have already been prototyped in Workpackage 4 consonant with these recommendations. Similarly, Chapter 7 discusses how the network traffic and user-movement analysis of Chapter 5 of this deliverable motivate work on modelling collective behaviour in Workpackage 5 in ways which in turn permit an integration of Workpackage 5 with the event design notations being worked on in Workpackage 4. In short, our concerted work on *OOTW* has levered new relations between all of the currently active workpackages in the project enabling some cross-workpackage, cross-partner and cross-profession collaborations which existed only to a lesser extent in the project before.

0.6 Dissemination and Publication

Part of the essence of eRENA is that the shows, installations, performances, exhibitions and broadcasts that the project produces are public events, accessible in principle to all. In short, a great part of our dissemination activity in the project is through the demonstrators, all of which publicise their relation to the project, to i3 and to ESPRIT. In this, *OOTW* was no exception, the fact that it was a product of funded European research being made clear in publicity surrounding the event, in discussions with the audience and in press reviews, most notably in the UK's national newspaper *The Times*.

Furthermore, attempts to disseminate the research work described here through prestigious academic conferences have already been initiated. A paper detailing the camera control and event management applications, together with ethnographic evaluation of them, has been submitted to SIGGRAPH99. A paper describing *OOTW* in general has been submitted to ECSCW99, the European Conference on Computer Supported Cooperative Work. Both of these papers have involved co-authorship between the computer scientists and industrialists, the social scientist and TV producer involved with *OOTW* and who also contribute to this deliverable. A third paper, concentrating purely on the ethnography of *OOTW* (cf. Chapter 6 of this deliverable) has also been submitted to ECSCW99.

Chapter One: Introduction
Inhabited Television—broadcasting interaction from within collaborative virtual environments

Steve Benford, Chris Greenhalgh and Mike Craven
Department of Computer Science, The University of Nottingham, Nottingham, UK

Graham Walker, Tim Regan and Jason Morphett
BT Laboratories, Martlesham Heath, UK

John Wyver
Illuminations Television, London, UK

1.1 Introduction

In this chapter, we introduce the idea of *Inhabited Television*, the combination of collaborative virtual environments (CVEs) and broadcast TV to create a new medium for entertainment and social communication. The defining feature of this medium is that an on-line audience can socially participate in a show that is staged within a shared virtual world. The producer defines a framework, but it is the audience interaction and participation that brings it to life. A broadcast stream is then mixed from the action within the virtual world and transmitted to a conventional viewing audience, either as a live event or sometime later as edited highlights.

Inhabited TV extends traditional broadcast TV and more recent interactive TV by enabling social interaction among participants and by offering them new forms of control over narrative structure (e.g., navigation within a virtual world) and greater interaction with content (e.g., direct manipulation of props and sets). Inhabited TV also builds on recent research into CVEs as social environments, such as experiments with Internet-based virtual worlds (Damer, 1997). In particular, Inhabited TV demands a more explicit focus on issues of production, management, format and participation arising from the staging of events within virtual worlds. The potential size of Inhabited TV audiences also challenges the scaleability of CVEs and in particular, the application of different awareness management and scoping techniques.

Our argument unfolds in two parts. In this chapter we present three early experiments with Inhabited TV: *NOWninety6*, a public poetry performance using a CVE; *The Mirror*, a series of on-line virtual worlds that ran in parallel with a conventional TV series; and *Heaven & Hell – Live*, a live TV broadcast from within a CVE. These experiments raised some fundamental questions for Inhabited TV and framed our subsequent research agenda. How could we engage on-line performers with an on-line audience? How could we establish fast-pace scripted interaction involving members of the public using a CVE? How could we produce a coherent broadcast from the action in the CVE? Finally, could the experience be engaging, both as a participant in the CVE and as a viewer of the broadcast?

The following chapter in this Deliverable presents our reaction to these questions in the form of a further experiment called *Out Of This World*. The defining features of *Out Of This World* were the formulation of Inhabited TV design principles intended to encourage coherence and social interaction, combined with the use of dedicated production software. The latter (see Chapter 3) enabled the show designers to configure the temporal structure of the show as part of the CVE and then provided interfaces for directing the show and controlling virtual cameras within this structure.

Before presenting our various experiments, we first introduce the idea of layered participation as a mechanism for describing Inhabited TV applications and for defining associated terminology.

1.2 Layered Participation in Inhabited TV

Inhabited TV can be described in terms of three layers of participation as shown in Figure 1.1. Each layer corresponds to the use of a distinct combination of interface and network technologies to access the shared virtual world and its contents, and therefore defines different possibilities for navigation, interaction, mutual awareness and communication. This division into layers is motivated by the need to structure the content of an Inhabited TV show, to map Inhabited TV onto a range of available delivery and access technologies, and to scale to large numbers of participants. To these visible layers may be added the additional invisible layer of production, those participants whose job it is to facilitate the show.

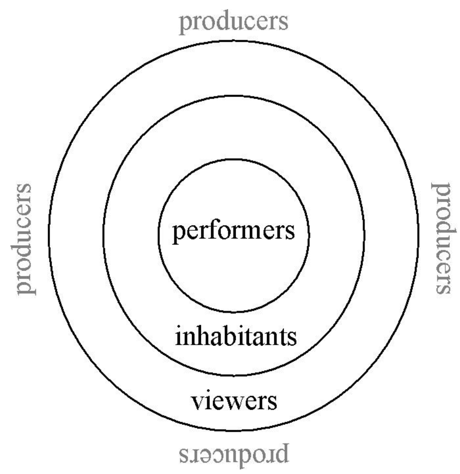

Figure 1.1: Layered Participation in Inhabited TV

1.2.1 Performers

The innermost layer describes the performers in the TV show (e.g., hosts and actors). These typically have the fullest involvement in the show requiring the greatest commitment and the richest forms of expression. In turn, this may require the support of relatively powerful equipment such as immersive peripherals, high performance workstations and high-speed networks. Performers represent core content and typically have global visibility in terms of being seen by the other layers. As each performer's data has to be broadcast to all other participants, the number of performers will be limited by available network bandwidth and processing power.

1.2.2 Inhabitants

The next layer describes the inhabitants, on-line members of the public who are able to navigate the virtual world, interact with its contents and communicate with one another. Inhabitants may have various kinds of involvement in a show including being an on-line audience (e.g., spectators at an event or a 'studio' audience), contributing content through some collective action or just socially watching the show in each other's company. Inhabitants typically use commonly available equipment. Currently this would be a commodity PC with an Internet connection, although in the future this may evolve towards a set-top box with access to a broadband public network. Inhabitants generally have limited visibility to the other layers within the content of the show. In the interests of scaleability, this may not be continuous global visibility on an individual basis (i.e., the detailed actions of every inhabitant need not be broadcast to all participants). Instead, individual inhabitants may only be visible to one another within some smaller sub-group and groups of inhabitants may be represented globally through some kind of low cost abstraction (e.g., collectively as a 'crowd', Benford et al. 1997a).

1.2.3 Viewers

The outermost layer describes the viewers who experience the show via broadcast or interactive TV. Viewers typically have only very limited possibilities for navigation and interaction. In the simplest case, they will be traditional TV viewers, i.e., the recipients of a broadcast mix that has been produced on their behalf and that can be received on a conventional TV set. However, interactive TV might offer them some additional possibilities such as choosing from among different perspectives or voting as part of large-scale audience feedback. In general, viewers are not visible within the content of the show (other than through abstractions of voting and similar feedback mechanisms). However, they may still contribute to an Inhabited TV performance and even be socially active via off-line feedback and discussion mechanisms.

1.2.4. Producers

Our final layer of participation describes the producers of an Inhabited TV show. In this case, production spans all aspects of technical support and 'behind the scenes' activity that may support a show. Examples include, directors, virtual camera operators, network engineers and software and hardware support. The producers may often be invisible to the other layers, although there may be exceptions, such as making virtual camera people directly visible to performers so that they can target their actions for viewers to see.

1.3 Early Experiences with Inhabited TV

We now describe three early examples of Inhabited TV in terms of our model of layered participation.

1.3.1 The NOWninety6 Poetry Performance

In November 1996 we staged a public poetry performance using the MASSIVE-2 CVE as part of Nottingham's *NOWninety6* arts festival (Benford et al., 1997b). The participation structure of this performance (using the terminology of the layered participation model from the previous section) is shown in Figure 1.2.

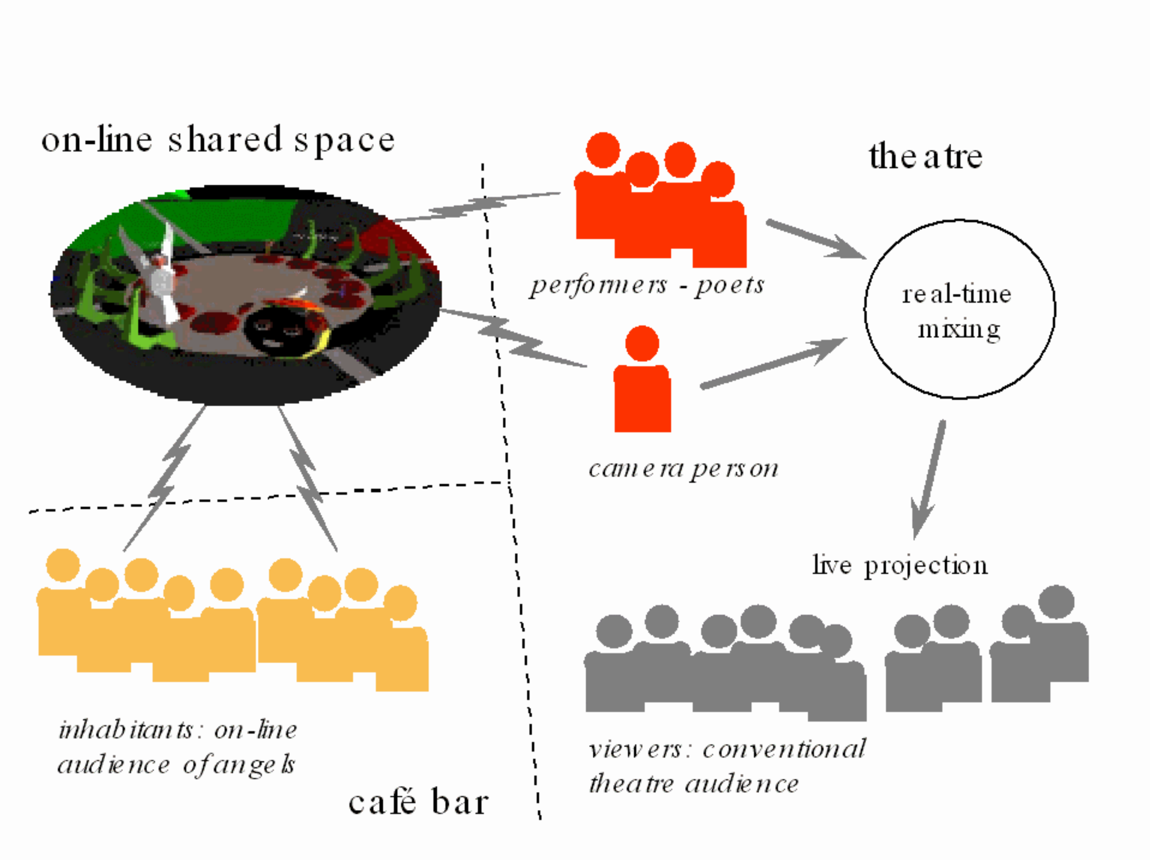

Figure 1.2: participation structure of NOWninety6.

The performers and viewers were co-located in a physical theatre. The performers were drawn from a company of 'hip-hop' poets and performed one at a time. Each poet appeared simultaneously on a stage in the physical theatre and on a corresponding virtual stage in the virtual world. Polhemus trackers were used to connect the poets to their avatars (at the head and two hands) so that their physical movements were directly mapped onto their virtual representations, thereby providing their avatars with additional expressive capability.

A simulated broadcast stream was mixed down in real time from the viewpoint of a virtual camera-operator (an example of a producer) who was located in the virtual world. This was projected into the theatre alongside each poet for the viewers to see.

Ten members of the public at a time could enter the virtual world as inhabitants using workstations that were located in a nearby café bar: Silicon Graphics O2 workstations were linked by an Ethernet. These inhabitants could move about, experience the poetry and could communicate with one another using real-time audio. During breaks in the performance, they were encouraged to explore four outer worlds that contained fragments of text from the poems. The inhabitants were therefore placed into the role of an on-line audience.

The event lasted for one evening and involved several cycles of performance and exploration. Approximately 200 people were in attendance of whom 60 experienced the virtual world in cycles of 10 at a time, with the remainder watching the 'broadcast' in the theatre. The screenshot in Figure 1.3 shows a performer (poet) avatar and several inhabitant avatars ('angels') near the virtual stage.

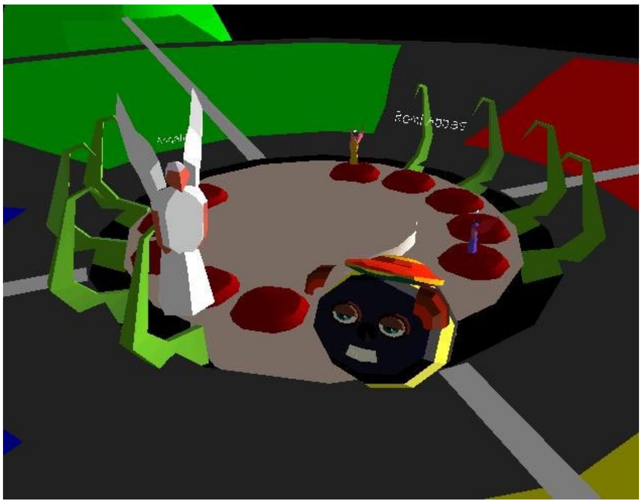

Figure 1.3: A poet and angels near the stage in NOWninety6.

1.3.2 The Mirror

The Mirror was an experiment in the first quarter of 1997 that involved public access to a series of six virtual worlds on the Internet. The experiment ran in parallel to the BBC television series *The Net* and the content of the conventional TV programmes provided inspiration for the design and content of the virtual worlds. Edited highlights of the action from the virtual worlds were shown on subsequent TV shows. Figure 1.4 summarises the participation structure of *The Mirror*.

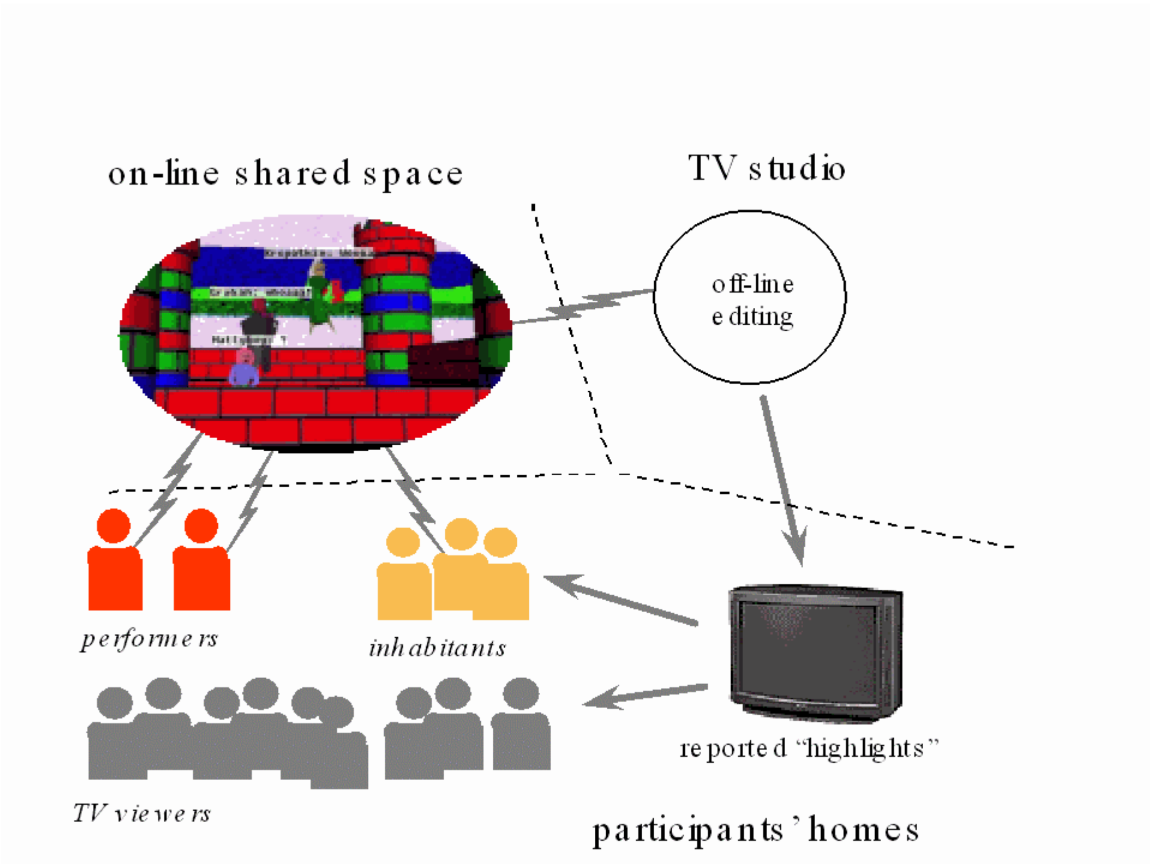

Figure 1.4: participation structure of The Mirror

The inhabitants (or 'citizens') accessed *The Mirror* from their homes or workplaces over the Internet. Sequences from within the worlds were recorded to videotape and edited for subsequent broadcast on TV, and hence inhabitants could become viewers at a later time.

The design of the six worlds of *The Mirror* reflected the content of the corresponding TV shows. The six worlds were:

- *Space* – a lunar terrain, populated by responsive aliens, which focused on issues of navigation.
- *Power* – a hall of fame featuring animated figures from the past and present of computing. This led to a debating chamber with audience voting facilities.
- *Play* – filled with games and tricks, designed to promote co-operation and competition. Examples included a shuffleboard game with persistent scores, a rocket that required three people to launch it and a bouncy castle.
- *Identity* – experimenting with notions of identity and the influence of the environment on people.
- *Memory* – a winding memory lane that passed through significant events from the last few decades. President Kennedy's motorcade would appear. Elvis

made fleeting appearances and a large clock displayed a count down to the 'end of the world'. Figure 1.5 shows an example scene from Memory.

- *Creation* – a world of vibrant flora and fauna which included an art gallery where citizens could display their own VRML 2.0 creations.

These six worlds were linked by an entry portal that highlighted a 'world of the week', corresponding to that week's broadcast TV programme. Various special events were also held within *The Mirror*, including: debates (e.g., between the science-fiction author Douglas Adams and Peter Cochrane, Head of Research at BT); a game show; an art exhibition; and the end of the world party.

Figure 1.5: A scene from Memory in The Mirror

The software used for *The Mirror* was *Sony's Community Place* (Lea et al., 1997). This supports text and graphical (but not audio) communication between inhabitants and performers. The minimum specification of equipment required to access *The Mirror* was a Pentium P90 PC running Windows'95 and a modem.

Over 2300 people registered to become citizens of *The Mirror* and received a CD-ROM containing the browser software and VRML2.0 content. Throughout the series, citizens spent over 4500 hours logged on to the server.

1.3.3 Heaven & Hell – Live

Our third example, *Heaven & Hell – Live*, was an hour-long game show that was staged inside a CVE and simultaneously broadcast live on the UK's *Channel 4* TV in August 1997. In other words, access by inhabitants and broadcast to viewers happened simultaneously, with the latter seeing the activities of the former.

The overall participation structure of *Heaven & Hell – Live* is shown in Figure 1.6. The performers consisted of a host and two contestants (all celebrities on UK TV) as well as three 'reporters' who provided additional commentary on the activities of the inhabitants (or 'lost souls' as they were called). The performers were located in an Inhabited TV studio along with the production team who were responsible for creating the live TV broadcast. This studio combined a local network of PCs, a TV outside broadcast unit and a high bandwidth Internet connection. The production team included director, vision mixer, sound mixer and production assistant. They had access to six virtual cameras within the world, taken from the viewpoints of the host, contestants and reporters, with the latter responding to instructions from the director.

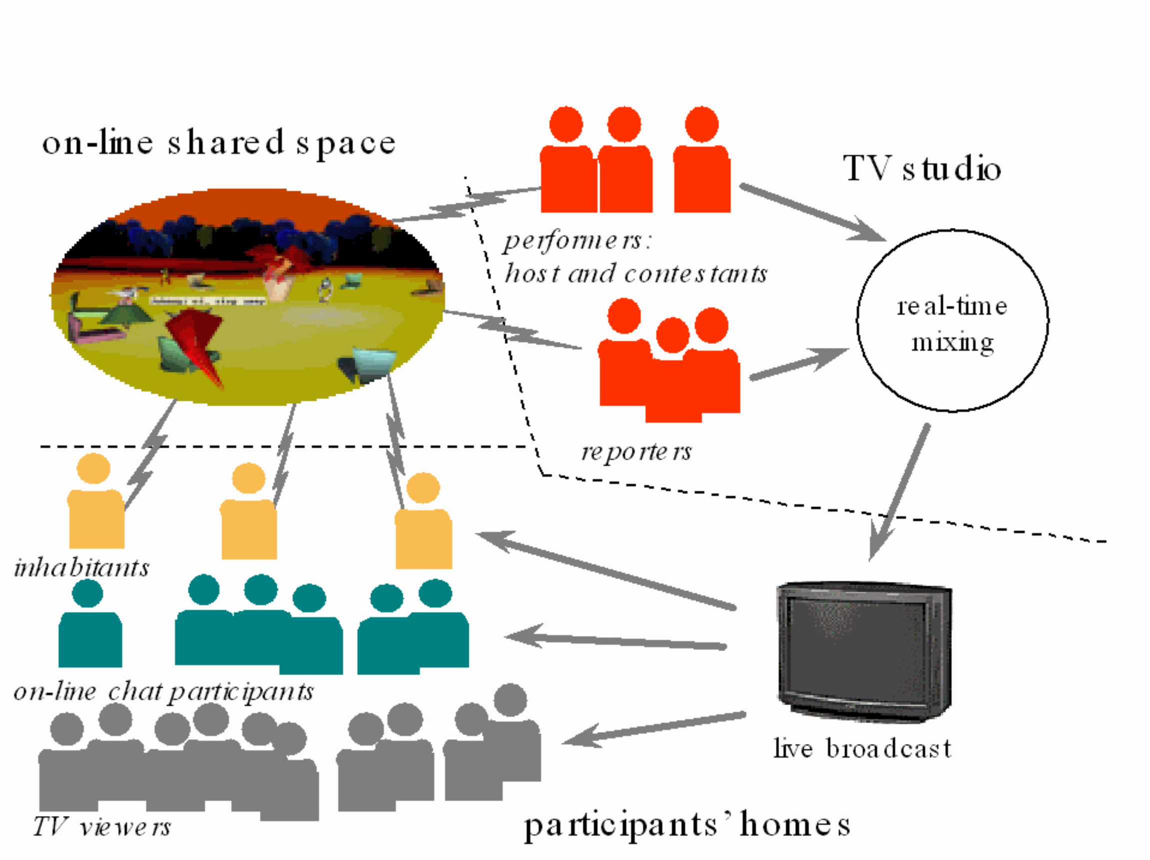

Figure 1.6: participation structure of Heaven & Hell - Live

As with *The Mirror*, the inhabitants accessed the shared virtual world from their homes via PCs and modems. The live TV broadcast was also available to viewers within their homes, and it was possible for inhabitants to also be viewers - i.e., to be watching the TV broadcast while logged in to the virtual world. Indeed, informal enquiries suggested that there was often a small group of viewers with a single PC and TV in the same room.

It is important to note the different modes of communication that were enabled in *Heaven & Hell – Live*. Like *The Mirror*, this event was implemented in Community Place and so communication within the virtual world (i.e. among performers and inhabitants) was via text and graphics. In addition, a live audio stream containing additional spoken communication between the performers was added to the final TV broadcast. This audio stream could not be heard within the virtual world, although its content was loosely reflected in the text chat. However, any inhabitant who was also watching the show on TV would have heard the audio. As an additional note, in order to support this complexity, the performer avatars were actually driven by two people: an assistant to navigate and type the text messages and the celebrities, who focused on the loosely scripted audio in the studio.

Games within *Heaven & Hell – Live* included a participatory treasure hunt through the world, a pursuit quiz and a gambling game, in which numbers of lost souls were wagered against random events revealed by the opening of 'evil-looking' pods in the 'bowels of hell'. Each of the two competitors was meant to collect lost souls as they went along, and the intention was that the inhabitants themselves were able to assist or hinder with the various games and tasks. Figure 1.7 shows a scene from the pursuit quiz part of the show.

The program was broadcast in August 1997 in a late-night slot. The on-line audience of inhabitants peaked at 135. In order to manage this relatively large number of simultaneous inhabitants the Community Place *aura* mechanism was used (Lea et al., 1997). This dynamically grouped the inhabitants into so called aura-groups of eight mutually aware participants. These participants could see and communicate with one another via the text chat and could also see the performers who were globally visible. However, the members of a group could not see the members of other groups. The same was true of the reporters (or virtual cameras) that at any moment in time would be seeing the performers and only one aura group. The viewing audience was estimated at 200,000.

Figure 1.7: A scene from Heaven & Hell - Live

This concludes our introduction to early experiments in Inhabited TV. The key characteristics of the participation structure of each experiment are summarised in Table 1.1. The following section discusses the lessons that have been learned from these experiences.

	NOWninety6	*The Mirror*	*Heaven & Hell – Live*
performers	Six poets performing sequentially over a two-hour show. Audio and body tracking.	Celebrity "super avatars" introduced for a small number of special events – debates and a game show. Text chat and interaction capabilities similar to the inhabitants.	Celebrity host and two game-show contestants. Audio for the TV broadcast, and assistants to provide text, navigation and interaction in the on-line world. Also three reporters with TV broadcast audio in addition to on-line interactivity.
inhabitants	Sixty, participating in groups of ten "angels" from locally situated silicon graphics workstations. Audio chat capability.	Over two thousand registered "citizens", recording 4,500 on-line hours over a seven week period from their home PC's. Text chat, shared behaviours and collaborative tasks.	135 "lost souls", participating in the one-hour live broadcast from their home PC's. Text chat and collaborative tasks. Also 800 users of the associated website chat and newsgroup during the broadcast.
viewers	Two hundred at a live performance, with the virtual world simultaneously projected into the cinema.	Half a million viewers of the BBC2 series "The Net" saw edited highlights and reporting on the worlds.	An estimated, two hundred thousand viewers of the Channel 4 TV broadcast.

Table 1.1: Layered participation in experimental Inhabited TV events

1.4 Lessons Learned and Issues Raised

The motivations behind our early experiments in Inhabited TV were to gain insights into its feasibility as a medium for entertainment, education and communication, and to help inform the design of future Inhabited TV content and technology. This section identifies the key lessons that were learned from these experiments. Our observations stem from two sources. First, they include the opinions of the inhabitants and viewers themselves as voiced over email for *The Mirror* and *Heaven & Hell – Live* and at a post-performance debate for *NOWninety6*. Second, they have emerged from a series of post-event group discussions. The lessons learned can be grouped under two headings: problems with establishing coherent social interaction within a CVE and problems with producing a coherent broadcast output from a CVE.

A third important area, problems in technical delivery and service management of a CVE, is not explicitly addressed in this Deliverable. This area includes issues such as security, authentication, content distribution, support and governance. These were most immediately apparent in *The Mirror* and *Heaven and Hell – Live,* where the inhabitants were truly distributed and using their own PC's, and the output was genuinely for broadcast TV. They were therefore not addressed by *Out Of This World*, which as we shall describe later, was conducted in a more controlled setting, similar to the arrangements for *NOWninety6*.

1.4.1 Establishing Coherent Social Interaction Within a CVE

One of the goals of Inhabited TV is that viewers will become more involved in a TV show by becoming inhabitants – they will become socially active and will contribute to a show in various ways. However, our early experiences suggest that it is difficult to engage members of the public in a coherent, real-time and fast-pace narrative within a CVE. We break this issue down into four further issues: problems with engagement between performers and inhabitants; difficulties with precise and coordinated movement; the difference between the pace of interaction in CVEs and that of TV; and the use of awareness management techniques.

Lack of engagement between performers and inhabitants

It proved difficult to establish a productive engagement between the inhabitants and the performers in all three events. At one extreme, the inhabitants were unable to get a word in edgeways as the performers dominated the interaction. At the other, the inhabitants spent all of their time chatting to one another and ignored the performers.

In *Heaven & Hell – Live* it had been planned that the inhabitants would interact with performers during several of the games (e.g., building stacks of avatars, suggesting answers in the quiz and helping with the treasure hunt). However, with a few exceptions, these interactions generally failed to materialise to a significant degree. Instead, the show mainly unfolded among the performers, with the inhabitants mostly relegated to the role of additional scenery that could occasionally be commented upon.

In contrast, the inhabitants of *NOWninety6* were provided with audio communication. The problem here was that their audio dominated that of the performers to such an extent that the performance was lost among the general mêlée. There was little evidence of any self-regulation of the inhabitants' behaviour as might have been expected at a conventional poetry performance (i.e., where people can shout out if they wish to, but tend to refrain from so doing). It seems that the performers were unable to engage the audience or to exert any social pressure on them. There is also a question as to the extent to which the performers were interested in the inhabitants given the presence of the local theatre audience.

The Mirror provided the most successful examples of on-line social interaction, especially the emergence of some collaborative games (in fact, the idea of avatar stacking first emerged in *The Mirror*). It is notable that *The Mirror* was the one example in which on-line interaction was divorced from the requirement to *simultaneously* address an audience of viewers. Indeed, the whole idea of performers and scripted content was generally less well developed in *The Mirror* and the social interaction that emerged was mostly among the inhabitants. *The Mirror* was therefore more like an on-line community in a CVE that was seeded and strengthened through a parallel TV show, than it was an Inhabited TV show in its own right.

Achieving precise and co-ordinated movement within a CVE

A conventional TV show will require participants to move to precise locations at particular instants (e.g., standing on a mark so as to be in shot) and for several participants to move in a co-ordinated way. Inhabitants generally experienced difficulties achieving such precise and co-ordinated movements in a CVE. For example, in *Heaven and Hell – Live* the inhabitants experienced difficulties keeping up with the action as it moved between the three areas of the virtual world (Heaven, Hell and Purgatory) and found it hard to quickly co-ordinate their actions in games such as avatar stacking. The *NOWninety6* poetry event introduced two techniques to support movement. The first was setting a home position at the virtual stage for each inhabitant. The inhabitant could easily return to this position by using a single button on the interface. The second was the use of 'travellators' (moving walkways) to move inhabitants from one part of the world to another (e.g., from the central stage to the four outer worlds). These techniques were partially effective, although precise navigation was still difficult, and indicate some possible technical solutions to this problem.

The difference between the pace of CVEs and that of broadcast TV

The issue of movement is indicative of a broader mismatch between the pace of TV and that of CVEs. Conventional live TV shows are scripted, highly structured and have a fast pace that involves precise (i.e., to the second) timings for events (some of the reasons for this are discussed in Chapter 6). Conventional TV also relies primarily on video and audio information whose presentation may involve a series of rapid transitions produced by camera movements and cuts between cameras. In contrast, CVEs traditionally have a much slower pace, especially where text communication is used instead of audio. Even when audio is used, the pace of social interaction may be slower due to the need to navigate through the world and due to network delays and other technical constraints.

This difference in pace was most evident in *Heaven & Hell – Live* where the inhabitants were navigating a graphical CVE and were using text to communicate, but where the viewers were receiving a combination of graphics, audio and video. The audio and video were added to the live broadcast output and included a very fast-pace audio commentary from the performers and also video views of two of the performers inset into the graphics from the virtual world. A combination of delays (broadcast, interaction, navigation, local PC performance and Internet communication) meant that it took between five and ten seconds for inhabitants to visibly react to events in the show (there was no deliberate delay). This delay was too long for communicating with the performers as part of a fast-pace TV show and there were several examples of performers addressing inhabitants, but not waiting for their response. As a result, the inhabitants became detached from the action of the show. The performers' impatience is understandable—the pressure to produce a fast-pace broadcast took precedence over the need to accommodate the slower pace of interaction in the CVE.

Providing the inhabitants with real-time audio communication may have helped, although, it would have challenged the use of public Internet connections. However, we suspect that there would still be significant delays and differences in pace between broadcast and on-line action. In short, addressing the difference in pace required by viewers and inhabitants involves a difficult balancing act.

The use of awareness management techniques

Our next issue is quite different and concerns the use of awareness management techniques. These are widely employed by CVEs to increase the number of simultaneous participants. They typically involve dynamically limiting a participant's knowledge of a virtual world to be some appropriate subset of the whole. Recent examples can be found in the *NPSNET* system (Macedonia et al., 1995), which tiles a virtual world with hexagonal cells, so that each participant sends information only to their current local cell and receives only from this cell and its immediate neighbours. The *Spline* system (Barrus and Anderson, 1996) composes a world from arbitrarily shaped regions or 'locales' which localise interaction and which may be 'stitched' together by various 3D transformations. In a different vein, *RING* (Funkhouser, 1996) scopes interaction and communication according to potential visibility in densely occluded environments while localising interactions at servers responsible for specific regions of the world.

In our own work, we have built upon the awareness management mechanisms of aura, focus and nimbus as originally defined in the spatial model of interaction (Greenhalgh and Benford. 1995). Aura is used to scope personal presence so that inter-personal communication is only enabled when participants' auras have collided. The Community Place software that was used for *The Mirror* and *Heaven and Hell - Live* realises a version of aura. This groups participants into aura groups using a *k*-nearest neighbours algorithm. This ensures that local aura groups have a guaranteed maximum size, thus ensuring that network traffic remains within well defined limits (a major consideration when supporting modem users) (Lea et al., 1997). MASSIVE-2 that was used for *NOWninety6* realises a version of aura based upon the detection of collisions between personal sub-spaces that may have different sizes for different individuals and different media. MASSIVE-2 also supports the mechanisms of focus and nimbus for controlling the way in which information (graphics, text and audio) is perceived and projected respectively by each participant. Finally, MASSIVE-2 includes a notion of boundaries with different degrees of visibility and audibility (for extensive discussion of this feature of MASSIVE-2 and of the use of different kinds of boundaries, see Deliverable D7b.1).

Heaven & Hell – Live used Community Place auras to divide the inhabitants into groups of eight mutually aware avatars who could see one another and could exchange text messages. The performers were represented by globally visible super-avatars. At any moment, a participant could see the globally visible performers and the seven other avatars within their current aura group. This also applied to the virtual cameras that were generating the broadcast output; these would also see the performers and seven inhabitants at a time, even when 135 inhabitants were on-line. Although successfully and necessarily limiting network traffic, this use of auras resulted in considerable confusion. For example, inhabitants and viewers doubted that so many people were really on-line, as only a few could be seen on TV. Confusion also arose when inhabitants who were also watching the TV broadcast realised that they were in camera shot, but couldn't see themselves.

NOWninety6 used MASSIVE-2's aura, focus and nimbus to achieve different levels of visibility for the various participants. The problem here was that these needed to assume different settings as the event unfolded. When the poets were performing, a small inhabitant nimbus would have minimised interruptions (each inhabitant would have been heard only by their neighbours). This might have led to a more coherent on-line performance. On the other hand, when the inhabitants were exploring the four outer worlds during breaks in the poetry, a large nimbus would have helped them hear one another even at a large virtual distance and would have encouraged social interaction among them.

The use of awareness management techniques is necessary in the interests of scalability. However, their current realisations in Community Place and MASSIVE-2 needs to be more intuitive and dynamic. Their operation should be clear (e.g. be tied in with the structure of the world). Furthermore, it will be necessary to dynamically adjust them to support different kinds of participation at different stages of a TV show, just as one would normally control lighting or audio in a staged theatrical production. We suspect that these observations carry over to the other awareness management techniques described above.

1.4.2 Producing a Coherent Broadcast Output from a CVE

The second general issue raised by our early experiments is the quality of the broadcast output that is created from the action within the CVE. Our three experiments were quite different in this respect. The *NOWninety6* poetry performance involved projecting the broadcast output into a theatre alongside a conventional performance. *The Mirror* involved off-line editing of the CVE action to be broadcast on TV at a later date as edited highlights. *Heaven and Hell – Live* was the most demanding in that it involved producing an hour of live TV broadcast. Given the demands placed upon it, it is *Heaven and Hell – Live* that provides much of the focus for the following discussion. However, the other two experiments also offer some insights into the problems involved in creating a broadcast output from a CVE. We sub-divide this issue into three parts: camera control and navigation; audio control and expressive avatars.

Camera control and navigation

Camerawork is an essential part of conventional TV production. There are various forms of camera (e.g., boom and track mounted or handheld) and dedicated mixing facilities for editing a single broadcast stream from multiple cameras. Inhabited TV uses *virtual cameras* to capture the action within the CVE. The control of these virtual cameras was a significant issue in our early experiments.

Heaven and Hell – Live deployed six virtual cameras within the world. The output of these cameras was then fed to a conventional TV mixing desk. The cameras *in Heaven and Hell – Live* used the standard avatar control interface; they were essentially disembodied avatars. As a result, like the inhabitants, the cameras had difficulties in keeping up with the action. They even became lost on occasions and there were several moments when a view into empty virtual space was broadcast for a few seconds. Occlusion was also a problem, cameras were frequently blocked by inhabitants. When this happened, it was difficult for the cameras to move around the obstruction without loosing sight of the target.

The *NOWninety6* poetry performance was limited by using only a single camera, so that there was no possibility of cutting between different views. On the other hand, this camera was controlled using a dedicated interface based upon the principle of *object centred navigation.* The camera was constrained to move about a focal point (the performer) in such a way that it always faced this point. As a result the performer was never lost from view and it was easier to move around occluding objects without losing the target. This camera also allowed the operator to specify designated movements in advance that could then be smoothly animated by the camera software.

Audio control

Audio tends to be a neglected aspect of CVEs. Indeed, only one of our early experiments, *NOWninety6*, used audio for communication among inhabitants and performers. In this case, all participants could speak at the same time. Each participant received an individual audio mix that presented other participants' audio streams according to their positions in the world using a combination of attenuation with distance and stereo panning. However, beyond this, 'soundscaping' received relatively little attention. Audio quality was low (8-bit quality) and there was little production control over the mixing of audio from the virtual world into the broadcast mix. For example, it was not possible to dynamically pick out or suppress individual inhabitants.

Heaven and Hell – Live and *The Mirror* did not use audio for the inhabitants' communication. However, audio was used to provide ambient sound (e.g., background sound and music) and for spot effects (e.g., associating sounds with specific events in the world). In both cases, this audio was generated locally by each inhabitant's PC using samples on a CD-ROM that had been mailed out in advance. As discussed in section 3, *Heaven and Hell – Live* also used audio from the studio host, performers and reporters as an essential element in the broadcast mix.

In contrast, conventional TV provides extensive facilities for handling audio. These include the use of different kinds of microphone (e.g., boom mounted, stage mikes and tie-clips) dedicated systems and personnel for mixing and editing audio, both live and in post-production, as well as the extensive use of music and sound effects.

Expressive avatars

Our final issue concerns the expressiveness of avatars. The avatars in current CVEs tend to be limited in terms of their expressive capabilities. In many cases they are merely customised 3-D cursors that mark their owner's position within a virtual world. Although the expressiveness of avatars is of course, relevant to interaction within the CVE, we argue that it is a particular concern for the broadcast output. TV viewers are accustomed to a rich visual experience that includes trained human actors. Watching simple avatars, often with no facial expression or other gestures is likely to be unsatisfactory. It should be noted that our experiments did make some limited attempts to provide more visually interesting avatars. The performers in *NOWninety6* could gesture through an articulated avatar driven by the output of Polhemus magnetic trackers on their head and hands. The avatars in *Heaven and Hell – Live* used local computer animations to give some additional life to the performers on screen and to allow the inhabitants to trigger pre-programmed gestures. There is much on-going research into increasing the expressiveness of avatars, including work on humanoid modelling, gesture and facial expression (Capin et al., 1997). However, it is not clear at what point the computing and network infrastructure available in the home will allow these techniques to be used for inhabitants, although performers may be a different matter.

1.5 Summary

Our three experiments raised some fundamental challenges for Inhabited TV. We have grouped these under the general themes of establishing coherent social interaction within a CVE and producing a coherent TV broadcast from a CVE.

One might argue that the root of our problems lies in trying to integrate what ought to be two separate media – broadcast TV and CVEs, each of which has its own distinct characteristics. However, we propose that the convergence of computing and home-entertainment devices (e.g., PCs, TVs, games-consoles and set-top boxes) coupled with the parallel convergence of data networks and broadcast technologies, means that the distinction between these media will gradually blur, if not disappear. The current trend is one of diversification of, and inter-working between, different media types. Furthermore, we argue that these two media can benefit one another. CVEs can provide new forms of content for TV involving public participation in physically impossible scenarios, all at a low cost – a network of computers in an Inhabited TV studio should cost significantly less than the construction of physical sets within a conventional studio. In turn, TV can offer an impetus for social interaction with CVEs and may be a powerful tool for building and motivating on-line communities.

Our approach has therefore been to adopt the challenges raised in this section as a research agenda for Inhabited TV. As a result, the following chapter presents a fourth more recent experiment – *Out Of This World* – that attempted to address several of these issues by staging a fast-pace Inhabited TV show that was coherent, both in terms of interaction within a CVE and in terms of its broadcast output.

Chapter Two: A New Experiment in Inhabited Television – *Out Of This World*

Steve Benford, Chris Greenhalgh and Mike Craven
Department of Computer Science, The University of Nottingham, Nottingham, UK

Graham Walker, Tim Regan and Jason Morphett
BT Laboratories, Martlesham Heath, UK

John Wyver
Illumination Television, London, UK

John Bowers
Royal Institute of Technology, Stockholm, Sweden

2.1 Introduction

Out Of this World (*OOTW*) was a public experiment with Inhabited TV that was staged in front of a live theatre audience. The event was staged as part of *ISEA: Revolution*, a programme of exhibitions and cultural events that ran alongside *the 9th International Symposium on Electronic Art (ISEA'98*) held in Manchester and Liverpool in the UK in September 1998. There were four public performances of *OOTW* in the Green Room theatre in Manchester over the weekend of the 5th and 6th of September. These were preceded by two days of construction, testing and rehearsal.

OOTW was implemented using the MASSIVE-2 system, the same system that had supported the *NOWninety6* experiment described in the previous chapter. Useful features of MASSIVE-2 for *OOTW* included: support for up to fifteen mutually aware participants; streamed audio and video; immersive and desktop interfaces; and realising simple collaboration mechanisms using third party objects (Benford and Greenhalgh,1997) as described later.

Like *Heaven and Hell – Live*, *OOTW* was a game-show. This choice allowed a direct comparison to be made between the two experiments. Given the observations in Chapter 1, the design of *OOTW* was motivated by two key questions:

- could we involve members of the public in a fast-moving TV show within a collaborative virtual environment? In particular, could we clearly engage the inhabitants with the performers and with one another, could they keep up with the action, what would they contribute, and would they enjoy the experience?
- could we produce a coherent broadcast from the action within the CVE? In particular, would the broadcast output be recognisable as a form of TV and would it be entertaining to watch?

The remainder of this chapter provides a brief overview of *OOTW* in terms of layers of participation and the structure and content of the show.

2.2 Layered Participation in *OOTW*

We begin by describing how our layers of participation (see Chapter 1) were realised in *OOTW*. The overall participation structure of the show is summarised by Figure 2.1.

2.2.1 The inhabitants

OOTW adopted a 'cheesy' outer space theme. The inhabitants were divided into two teams, aliens and robots, who had to race across a doomed space station in order to reach the one remaining escape craft. On their way they had to compete in a series of interactive games and collaborative tasks in order to score points. The final game was a race in which these points were converted into a head-start for the leading team. The two teams each consisted of four inhabitants, members of the public who had been selected from the theatre audience. Every participant in the show could speak over a live audio channel. The teams were separated into women (aliens) versus men (robots) so that viewers would be able to more easily associate the voices that they heard with the avatars that they saw on the screen, although this turned out to be controversial decision as we shall see in Chapters 6 and 7. The team members were given cartoon like avatars that could be distinguished by a visible number on their backs and fronts. A speech bubble would appear above their heads whenever they were transmitting audio. The inhabitants used standard PCs with joysticks and combined headphone/microphone sets (see Figure 2.2). They were located behind the scenes, out of sight of the viewers in the theatre.

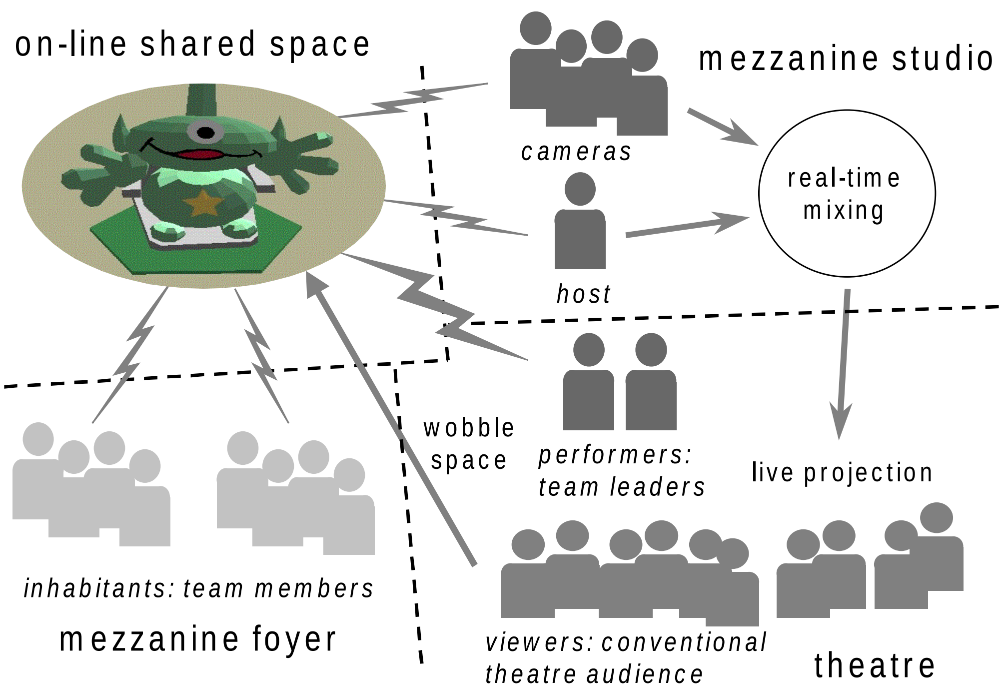

Figure 2.1: participation structure of OOTW

Figure 2.2: an OOTW inhabitant

2.2.2 The performers

The teams were guided by two performers, an actor and an actress, who played the role of team captains. The role of the captains was to encourage the teams to take part, to act as foci for the games and to improvise around the inhabitants' dialogue. The performers used immersive Virtual Reality equipment (Figure 2.3), including Polhemus magnetic sensors to track the positions of their head and both hands which were then represented on their avatars in order to give them a greater expressive ability than the inhabitants. Unlike the *NOWninety6* poetry performance, the performers were fully immersed (i.e. were wearing a head-mounted display). They were also given a virtual 'light stick' that they could activate by pressing a button on a hand-held flying-mouse and which allowed them to point at objects, locations and participants in the virtual world.

Although logically they would have been out of sight of any viewer at home, the performers were actually physically located in the theatre space so that the viewers could see them working with the immersive technology (Figure 2.4). This compromise was designed to enhance the viewers' understanding of the concept of Inhabited TV and to create a staging appropriate to a theatre space (see further discussion of this and related compromises in the design of *OOTW* in Chapter 6).

The show was hosted by a third performer who appeared in the form of a live video face that was texture mapped onto a large mobile virtual screen within the world. This screen would occasionally rotate around to show the game scoreboard.

Figure 2.3: An immersed performer

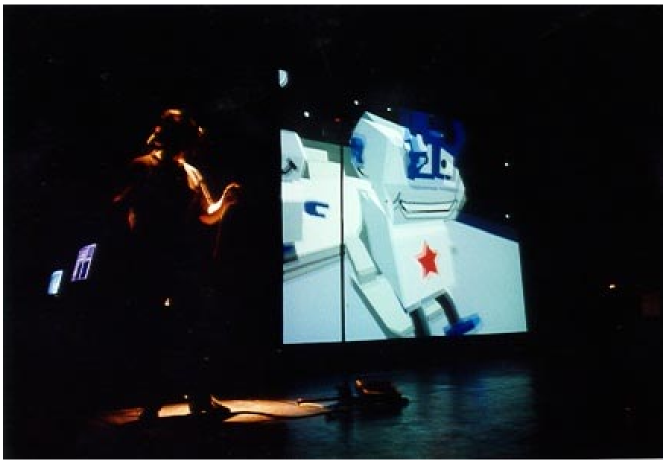

Figure 2.4: The performer located next to the screen in the theatre.

2.2.3 The viewers

The viewers were seated in a conventional theatre facing a large screen onto which the broadcast output was projected. The two performers were physically located on either side of this screen. For most of the time the viewers did not directly interact with the content of the show and as such resembled a traditional TV audience. However, they were provided with an opportunity for mass-interaction towards the end of the show. This involved them choosing the best losing team member through a mechanism called WobbleSpace (see Chapter 4). WobbleSpace was based on a demonstration of audience interaction using the CINEMATRIX interactive entertainment system

(CINEMATRIX, 1998) that was first demonstrated at SIGGRAPH. In *OOTW*, each member of the audience was asked to vigorously wave one of four coloured cards in order to express their vote. The overall level of activity of each colour was detected from a video image of the audience using image processing techniques and the resulting scores were passed to the CVE software. The audience was encouraged to test this voting mechanism by playing a game of 'Pong' in the pre-show warm-up in the style of the original SIGGRAPH event. The warm up also involved a brief explanation of the concepts behind *OOTW*. Finally, after the show, the viewers were invited to stay behind and provide us with feedback.

2.2.4 The production team

The show also involved an invisible but essential production team who were responsible for managing the CVE software and for producing the broadcast output. A detailed empirical description of the work of this team in realising *OOTW* forms the focus of Chapter 6; let us just give an overview of their work here.

Four virtual camera-people were present in the virtual world, although they were not visibly embodied. Using purpose built interfaces (see Chapter 3) running on graphics workstations, they were able to capture the action from various perspectives. Video and audio output from their computers was then fed into a conventional TV mixing desk where it was mixed by a professional TV director and her assistant. The resulting video mix was sent to the projector in the theatre. In addition, a world manager was able to control the virtual world software, including activating virtual objects and constraining the actions of the participants (see Chapter 3). Figure 2.5 shows the director and her assistant at the mixing desk. The four camera monitors with feeds from the virtual cameras can be seen on the far left. To the right of these is the current transmission monitor (showing the actual broadcast) and a monitor for previewing video material (the face of the host, the title sequences or other videotape inserts). Figure 2.6 shows the four virtual camera operators at their machines (physically located just behind the director).

Figure 2.5: The director and her assistant

Figure 2.6: The four virtual camera operators

2.3 The Structure and Content of *OOTW*

We now move on to consider the structure and content of the show. *OOTW* involved a journey through five arenas that were joined together into a linear structure by a series of virtual travellators. Each arena involved the two teams in a different task as follows:

- *Arena 1: introductions* – an overview of the show from the host followed by introductory statements from the captains and individual team members (Figure 2.7).
- *Arena 2: flipping frogs* – a collaborative action game in which the teams had to flip space-frogs onto spiky hats worn by their team leaders. Flipping was achieved by closely approaching a frog, causing it to jump away in the opposite direction. The teams had to impale the most frogs to win the game (Figure 2.8).
- *Arena 3: falling fish* – the team members had to harvest space-fish by collaboratively lifting their team leader up into the air and moving them about so that the leader could knock the fish from the ceiling by swiping them with their hands. The team leader was on a platform whose height varied according to the number of team members that were inside it and whose position was the average of its current members (Figure 2.9). The team members therefore collectively steered the platform and the team leader could only reach the fish when all four team members were inside. The team that harvested the most fish won the game. The platform is an example of a third party object in MASSIVE-2 that combines a group membership mechanism with a dynamically computed aggregate group representation (Benford and Greenhalgh, 1997).
- *Arena 4: culture quiz* – a quiz where the host asked the questions and the team members conferred to agree an answer that was then relayed through the captains. Each team had to answer questions about the opposing culture (i.e. robots about aliens and aliens about robots). A point was scored for each correct answer, resulting in the captain being raised up through a hoop that would start spinning, accompanied by a fanfare.

- *Arena 5: space-car race, WobbleSpace and the end of the world* – the final game was a race in which the teams had to steer a space-car along a twisted course in order to knock down a series of cones (Figure 2.10).The space-car was steered in an identical way to the platform from the falling-fish game, i.e., the team members controlled it through their collective movement. The team with the most points from the first three games was given a head start. The first team to cross the finish line won the show and was transferred to the space-craft ready for their escape. The losing team members were then asked to state a case for why they should be saved. Following this, the distant viewers voted for the best loser using WobbleSpace and this loser was then transferred to the ship (see Figures 2.11 and 2.12). The climax of the show was then the ship departing and the space-station imploding.

While journeying along the travellators between each arena, the teams (and hence viewers) were shown a pre-prepared video of the next game that appeared as a video–texture on the host's virtual screen. At the start and finish of each arena the host would encourage the team captains and team members to comment on their play up to that point. As a final detail, in addition to various sampled sound effects, synthesised sound was played to convey a sense of the environment of each of the arenas (e.g. mutated watery sounds during falling fish), the motion of the travellator, the take-off of the space-craft and the space-station imploding.

Given this general introduction to *OOTW*, the following section now focuses on the steps taken to address the lessons of previous experiences.

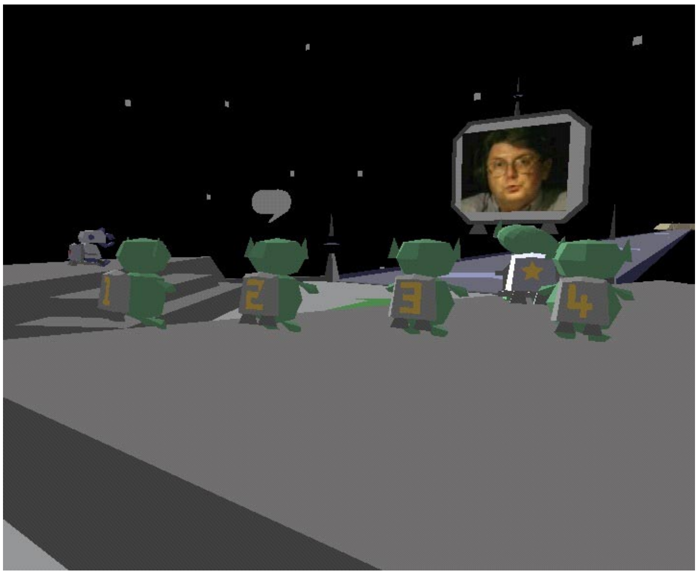

Figure 2.7: introducing the Alien team.

Figure 2.8: the alien captain (and hat) in flipping frogs.

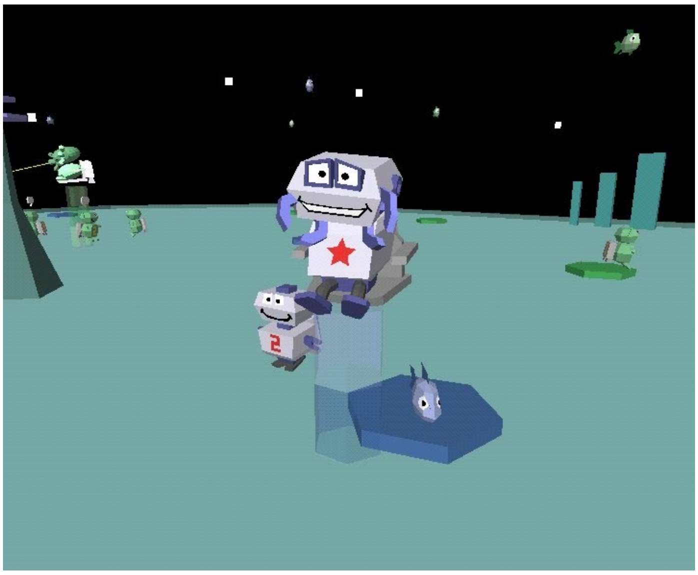

Figure 2.9: the robot captain on his platform in falling fish.

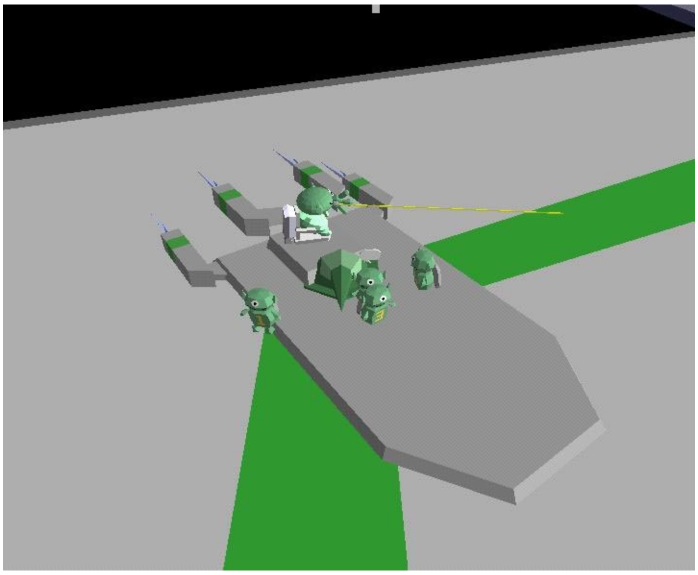

Figure 2.10: the aliens race their space car.

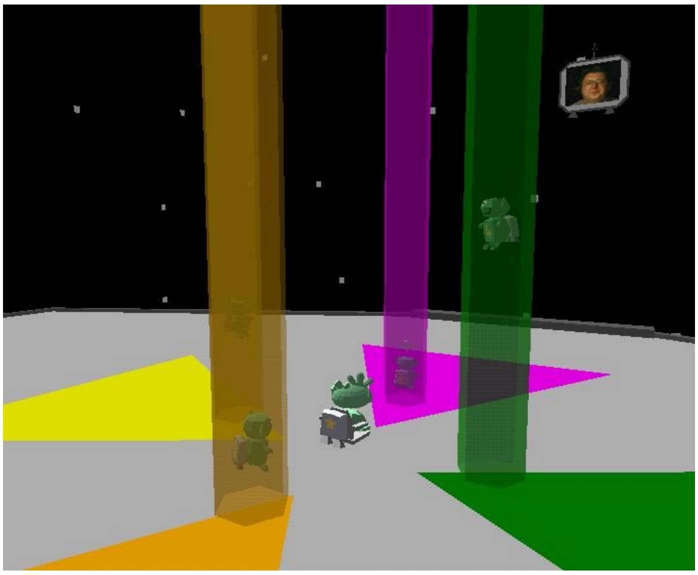

Figure 2.11: the aliens in WobbleSpace.

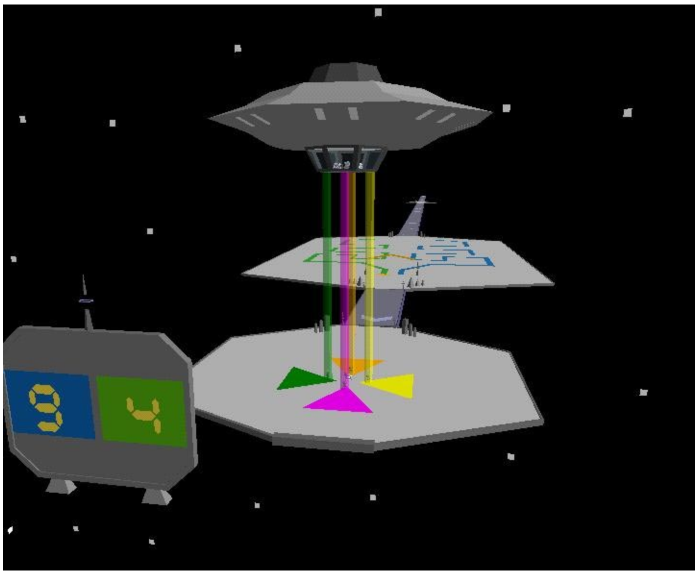

Figure 2.12: the end of the show.

2.4 Summary: Show Design Principles for *OOTW*

OOTW addressed the lessons from previous experiences through two innovations. First, we introduced a set of design principles, intended to increase the coherence of the show in terms of its visual appearance, social interaction and narrative structure. Second, we developed a suite of production software to support the temporal structuring of the show, the application of constraints on participants' actions and the control of virtual cameras. The following chapter gives details of the production software. Here, let us conclude by being explicit about the design principles which governed our work on *OOTW*.

In a direct response to our previous experiences with Inhabited TV, especially with *Heaven and Hell – Live*, the structure and content of *OOTW* was designed according to several key principles. These were intended to maximise the coherence of the interaction within the world and of the broadcast output and to establish a clear engagement between the participants, especially between the performers and inhabitants. Briefly stated, these principles were:

- **simplicity of concept and representation –** the games should be as simple as possible in terms of concept, interaction required and graphical representation. Emphasis was to be placed on interaction mechanisms rather than on graphical richness. In fact, the graphical designer was introduced at a relatively late stage once the entire show had been protoyped.
- **clear roles for participants –** the roles of the inhabitants and performers should be clearly defined and the outcome of each game should depend upon both (no-one should be relegated to the role of observer or mere 'helper').

- **co-operation** – the games must require co-operation, both between inhabitants and between inhabitants and the performers.
- **interaction through proximity –** we favoured indirect interaction with objects (e.g. having objects that react to the proximity or movements of participants) rather than direct manipulation of objects (e.g. selecting them with a mouse). This principle ensured that participants only had to learn to perform two tasks: moving about the world and speaking into a microphone. It also encouraged participants to engage in the highly visible and relatively interesting activity of moving about and required them to get close to the objects of interest as opposed to standing back and picking them off from a distance. In this way, it was intended that the composition of camera shots depicting close action could be facilitated.
- **action at ground level –** we generally restricted participants' movements to be at ground level. This was intended to simplify their movement as well as enhance camera work. For example, camera operators should be able to more easily locate participants than if they were distributed throughout 3D space. Furthermore, the vertical dimension would be free for pulling back and up to get overview shots.
- **playing to our technical strengths** – put simply this involved exploiting MASSIVE-2's most interesting features and avoiding its weaknesses. Its key features include potentially mobile aggregate group objects (Benford and Greenhalgh, 1997) and the general availability of audio communication. Its relative weaknesses include coping with graphical complexity and large numbers of animations and textures.

Chapter Three: Production Support Technology

Mike Craven, Steve Benford and Chris Greenhalgh
Department of Computer Science, The University of Nottingham, Nottingham, UK

3.1 Introduction

This chapter presents details of the production support technology developed for the *Out Of This World* demonstrator. This comprised the Event Management tool and Camera Controls. These are both enhancements to Nottingham's MASSIVE-2 CVE platform.

3.2 Event Management

Prior to *OOTW*, the CVE platform implemented by MASSIVE-2 allowed location and trajectories of all artifacts to be defined in a single world definition script, which are preset when the world server is initiated. When users joined such a world, they were free to navigate impeded only by the placing of solid boundaries, or by a movement field which if entered would move the user within it. Users were also completely free to manipulate artifacts, unless fixed, usually by picking them with the mouse. Thus there was no concept of external directorial control over objects and users once the world was started, and most significantly no way to control the pace of a show other than by relying on the participants.

In order to provide a degree of directorial control over both objects and users in the virtual world, involving imposed changes to locations and trajectories of objects and users at different times throughout the show, the idea of an event manager process was conceived. The key concept in event management in *OOTW* is the division of the game flow into *phases*. In a single phase, the location or trajectory of objects, user constraints, and other parameters, may be set or reset to facilitate progression of the show. This effectively means that the world is configured not just once, but many times throughout a show. User constraints are introduced to limit the extent of navigation of participants.

3.2.1 World Hierarchy and Object Positioning

Spatial configuration of a virtual world usually entails supplying location coordinates with reference to a global origin. In MASSIVE-2 a 3D position vector and orientation information in some form (rotation matrix or quaternion) is supplied for each artifact. MASSIVE-2 also allows the setting of trajectories, i.e. preset motion paths. However in a complex world with a large number of objects and users, which may be changing at every phase, this would become unmanageble unless some kind of world structure is imposed. We chose to overcome this in the first instance by enhancing world structuring with hierarchical scripting, allowing areas of the world to be defined with a local origin which are then included into a master world file. Related to this, a facility was added to allow objects or users to become the 'children' of others, such that the origin of the child object moves with the 'parent' object. Having a world hierarchy makes the design and maintenance of arenas much easier, including event management. Furthermore, it helps to make world geography much more readily understandable to the designer and also allows each arena to be set-up separately without worrying about its overall placement in the game-show world. We chose to make the local origin of each arena in the centre, so that objects and users within were positioned in relation to this. In addition, abstract objects were used to define named constraint regions for participants, called *user-hints*, and also for camera navigation, called *camera-hints*. Figure 3.1 shows a schematic top view of an arena with local co-ordinate system, an object ♥ at (10,0,20) and a constraint region at (0,0,-30) with user ☺ inside at (-10,0,-30).

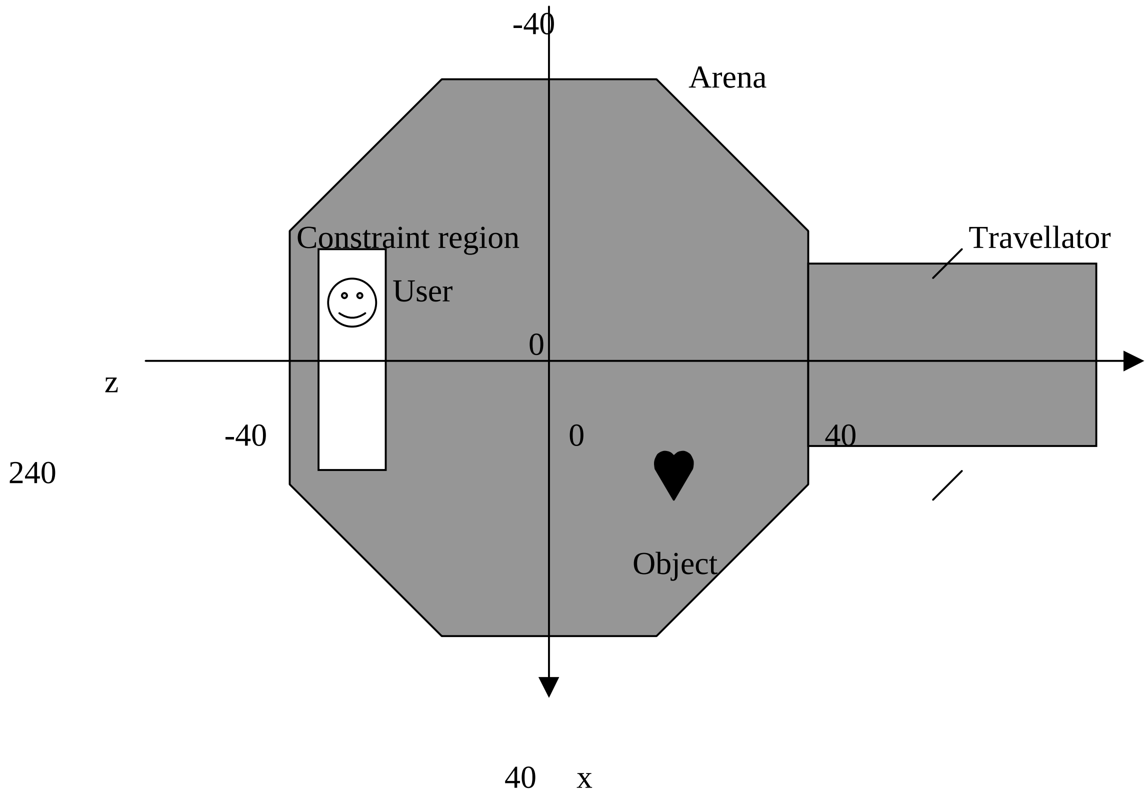

Figure 3.1: Arena and Travellator (viewed from above at y=0)

In order to add phase dependency to an object in the world, the following PHASE_INFO descriptions were added to the object definition, one set required for each phase in which the object appears:

- Phase identifier – an integer number, unique to that phase
- Reset flag – if TRUE, object is returned to a default location

- Appearance flag – if TRUE, the object appears in this phase, otherwise set null trajectory
- Trajectory – if defined, location or trajectory of the object is specified, otherwise previous trajectory is maintained

Constraint regions are defined as bounding boxes. This box might be small enough that a user can only turn on the spot, or it could be large enough for them to be able to explore a large area. When the world server process is invoked, the user-hint and camera-hints are extracted to separate files in global co-ordinates. Each hint has a text name (e.g. “arena1-start”) which is used to identify that hint. The use of camera hints is explained in a later section.

In addition to phase information, certain objects can also be given unique names so they can be identified by the event manager.

3.2.2 Event Manager Script

In addition to the phase specific information for objects, the designer is able to configure a number of properties for each phase which are defined in a separate file called *the event manager script.* This file ootw.mgr is read by the Event Manager after the world server is up and running (after the world hierarchy has been flattened).

Properties include:

- Phase identifier
- A text name for the phase
- Whether this is a roll-over phase, in which case the next phase is automatically activated after a specified time interval.
- Whether this phase can be selected manually
- Special flag indicating this a WobbleSpace (audience participation) phase
- User list – contains further properties for each named user or group affected by this phase. May also contains named objects.

Phase identifiers corresponds to those in the world definition files. Text names, supplemented with “R” if a rollover phase, appear in the event manager’s user interface.

Each entry in the user list contains the following properties:

- User or group name—this is the name of the user defined in the embodiment file e.g. “Inhabitant 1A”, or a group name which may be present in several embodiment files e.g. “Team A”. Previously named objects can also be specified here.
- Parent name—so that a user or named objects can be attached to another in this phase. For example, attaching a ‘costume’ to a participant.

- User constraints—these are specified as a starting constraint and a finishing constraint for this phase. The positions (in global coordinates) and extent of the regions can be given directly, or preferably the name of a user-hint is given corresponding to a predefined constraint region. When a phase is selected the user is immediately placed within the starting constraint region. If the finishing constraint region is different and a time period is also specified, the bounding box gradually changes its extent and position over that time until it corresponds to the finishing region. The user or members of the group are moved so they remain with the bounding box. While this is happening, and more generally whenever the participant is being constrained, optional feedback is given to the participant to inform them this is happening (by added a red filter to the console window). In other cases the boundary extent is just increased to give a larger region for the participant to navigate within.
- Audio levels and extents (as defined by audio nimbus in MASSIVE-2) for specific participants.
- Viewpoint—the participant’s viewpoint in the world can be selected from 1 of 6 choices predefined in the embodiment file, or defined manually.
- Highlighting—this allows the event manager to change the geometry of a participant in order to highlight them during this phase e.g. by adding an arrow over the participant’s head.

In addition to specifying new properties, properties can be left unchanged from the previous phase. If a participant or group is omitted from the user-list, they are left unchanged by that phase.

3.2.3 Event Manager User Interface and Script Construction

A dedicated event management user interface was developed to support a member of the production team in dynamically selecting different phases of the event as the show progressed. The phases were presented as a list and any phase could then be selected by name, causing the whole event to jump to that phase. By following a script (and taking cues from the game-show host), the event manager could push the show along, moving participants to their correct positions and initialising objects. In this way the show could be made to run to a strictly timed schedule and the participants could be brought together into a structured arrangement at key moments before being released again into a more exploratory or unconstrained activity.

The names of the 67 phases in *OOTW* are show in Figure 3.2, split below into 6 sections corresponding to the starting arena, the 4 game arenas and the final WobbleSpace arena.

```
"Initialise All..."
"Starting Positions in Start Arena"
"Freedom of Movement in Start Arena"
"Move All to Start Arena Exit"

R "Transfer All to Arena 1"
R "Remove Cover 1 and Reset Frogs"
"...continued"
"Positioning in Arena 1"
"Arena 1 Free Teams and Start Game "
"Move All to Arena 1 Exit"

R "Transfer to Arena 2"
R "Remove Cover 2"
"  Lower Fish for 4 in each team"
"  Lower Fish for 3 in each team"
"  Lower Fish for 2 in each team"
"  Lower Fish for 1 in each team"
R "Positioning in Arena 2"
R "(move captains to origin)"
"   (parent captains to plaforms)"
"Arena 2 Free Teams and Start Game"
"Move All to Arena 2 Exit"

R "Transfer All to Arena 3"
R "Remove Cover 3"
"Set Hoops"
"Positioning of All in Arena 3"
"Arena 3 Free Teams"
"Up Though Hoop 1 Alien Captain"
"Up Through Hoop 2 Alien Captain"
"Up Through Hoop 3 Alien Captain"
"Down Alien Captain"
"Up Though Hoop 1 Robot Captain"
"Up Through Hoop 2 Robot Captain"
"Up Through Hoop 3 Robot Captain"
"Down Robot Captain"
"Move All to Arena 3 Exit"

R "Transfer All to Arena 4"
"Remove Cover"
"Reset Cones and Jet Cars"
R "Position Captains in Arena 4"
R "  (move captains to jet-car origin)"
"   (parent captains to jet-cars)"
"Position Teams and Host in Arena 4"
"Arena 4 Aliens Start Race"
"Arena 4 Robots Start Race"
R "Move All to Arena 4 Exit"
"Replace Cover"

"Aliens Win - Transfer Aliens to
Spaceship"
"Robots Win - Transfer Robots to
Spaceship"
R "Aliens Win - Move Spaceship Up"
"Aliens Win - add tubes"
R "Robots Win - Move Spaceship Up"
"Robots Win - add tubes"
"Aliens Win - Move B to WobbleSpace
Position"
"Aliens Win - Do WobbleSpace for Robot
Team"
"Robots Win - Move A to WobbleSpace
position"
"Robots Win - Do WobbleSpace for Alien
Team"
"Aliens Win - Robot 1 Wins WobbleSpace"
"Aliens Win - Robot 2 Wins WobbleSpace"
"Aliens Win - Robot 3 Wins WobbleSpace"
"Aliens Win - Robot 4 Wins WobbleSpace"
"Robots Win - Alien 1 Wins WobbleSpace"
"Robots Win - Alien 2 Wins WobbleSpace"
"Robots Win - Alien 3 Wins WobbleSpace"
"Robots Win - Alien 4 Wins WobbleSpace"
"  Spaceship Moves Away and Host Leaves"
"  World ends!"
"  Spaceship and Host Leave"
```

Figure 3.2: Phase names and rollover indicators, as shown by the OOTW event manager interface.

As can be seen from Figure 3.2, the phase structure of a show can be quite fine-grained. The influence of phase management spans movement on the travellators, dialogue at the exit and entry points to the arenas and the structure of the games themselves. Examples of the use of phase properties in *OOTW* included moving the participants to start and end positions in each arena, moving them along travellators, attaching the team leaders to objects such as the spiky hats in the frog game, and resetting objects such as the spinning rings in the quiz game. In most cases the user-constraints keep the participants on the ground i.e. y=0. The main exception to this was in the quiz game where the captains were moved up diagonally through hoops.

In addition to these, two extra phase related events were devised for use on specific occasions in *OOTW*. The first of these was handling the audience participation from the WobbleSpace server. During a WobbleSpace phase, communication with the WobbleSpace server would be initiated, and votes from the audience's coloured card would be translated into an incremental height for each of the four WobbleSpace participants. Thus the participants would be raised up towards the spacecraft until one of them reached the required height. At the end of the phase, communication with the WobbleSpace server would be terminated and the losing participants would fall to the ground, the winner staying in position. The other special events were triggering of animation sequences which include the removal or replacing of the transparent 'domes' covering each arena, and the end of the world finale involving the imploding of the final arena and falling stars.

Phases were also used to represent branching points in the narrative, for example, choosing the next action according to which team had won a particular game. The manager could also choose to return to previous phases or to miss out phases. An example of repeating a phase was in the starting arena, where participants began the game in fixed starting positions, They were then given a few minutes to experiment with their controls and play football, before being moved back to their starting positions again. An example of missing out of phases was in the quiz game, where a captain only moves up through a hoop if their team answers a question correctly. Finally, several contingency phases were specified in the expectation that participant's equipment might fail. For example, there were versions of the falling-fish game in which the team leader could reach the fish if only three, two or even one team member was in the platform. The default case with four fish was automatically selected by roll-over from the previous phase, but this could be overridden by direct selection of the other versions.

Parenting of participants to objects within phases is seen in games 2 and 4 where captains are attached to the platforms and jet-cars, to be collectively moved around by the other participants. In fact this was the only way the captains could be moved other than by direct intervention by the event manager. In the WobbleSpace arena, the winning team members were parented to the spacecraft before it left the arena. After WobbleSpace, the additional participant from the losing team, chosen to be saved by the audience, must selected by the event manager to be parented to the spacecraft.

In order to repeat a game, some objects like the frogs must be replaced in their initial positions. This was achieved by resetting objects at the start of a scene.

To summarise, the event manager facilitates control over narrative in several ways:

- Divides the game-show story into phases, analogous to scenes in a play
- Supports direct selection of scenes
- Supports automatic rollover from one scene to the next
- Supports branching points, and multiple endings
- Supports multiple versions of scenes, including a default scene selected by rollover
- Supports repetition of scenes
- Provides for reset of objects within scenes

As well as the main event manager user interface, three other interface windows were provided (see Figure 3.3):

1) Audio output interface—to dynamically adjust participants' and effects audio levels

2) Direct control interface—dynamic control over highlighting, audio muting and movement
3) Score control—to update the in-world scoreboard

Figure 3.3 OOTW user interface windows

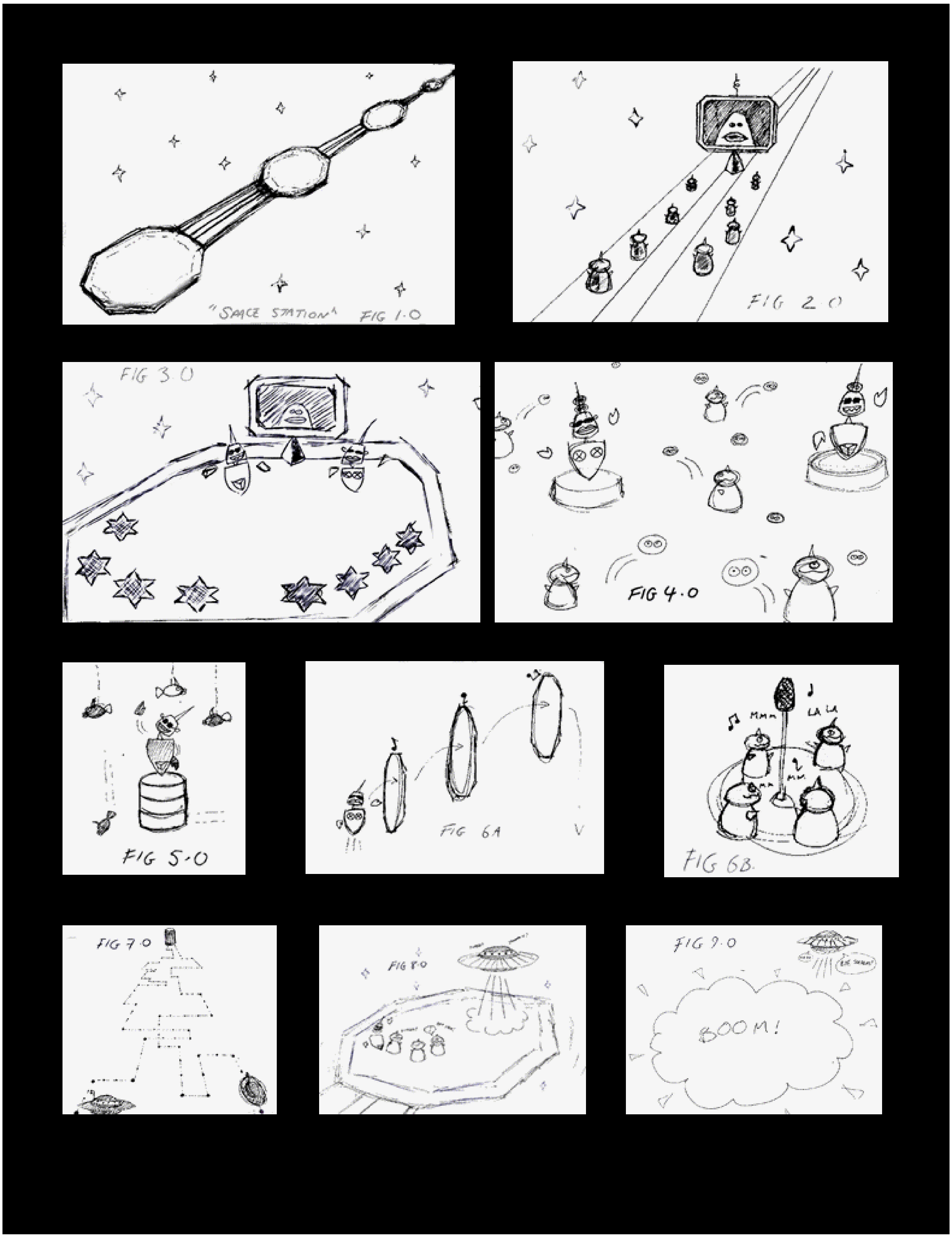

Figure 3.4 Storyboard for Out Of This World

3.2.4 Process: Storyboard to Event Manager Script

Starting from the *OOTW* storyboard it was necessary to break down the game show first into its higher level stages. As we were using arenas for the different parts of the game, this breakdown fell conveniently into 6 stages (analogous with acts in a play) taking place in separate arenas. Figure 3.4 showing drawings of the overall world structure, a travellator, and the content of the in the 6 arenas comprising:

1. Start Arena – introductions and preamble
2. Arena 1 – Flipping Frogs
3. Arena 2 – Falling Fish
4. Arena 3 - Quiz
5. Arena 4 – Jet-car Race
6. WobbleSpace Arena – audience participation, finale

Before describing the phases specific to each arena, a strategy was devised for users approaching and progressing within any game arena, under the control of the event manager. These were:

- Move all participants from the exit of the last arena to the start of the next, along the travellator
- Move participants to start positions (usually captains only)
- Release team members (to play game)
- Move all participants to exit

This progression was intended provide both director and participants a coherent understanding of how they should behave in the show.

Once the skeleton structure was devised, the next step was to identify all object and inhabitants to be controlled by the event manager.

All static (e.g. the arenas themselves) and behavioural objects (e.g. space frogs) were assigned PHASE_INFO descriptions which were added to the other object parameters contained in world definition files, one for each arena.

Inhabitants and all object which were involved in parent-child relationships with inhabitants were assigned unique text attributes e.g. “Inhabitant1” “A”, “Captain” “B”, “Platform” “A”. Prior to this all possible parent-child relationships were noted, e.g. captain-platform, spiky hat-captain, captain-jet car, inhabitant-spaceship so that the naming system could be devised.

Then, the precise offsets from the arena centres were determined for all user constraint regions, e.g. for entry and exit points, and abstract objects representing these were placed into world definition files, one for each arena. Each region was assigned a unique text attribute, e.g. “arena-1-entry-inhabitant1a”, “arena-3-pos-captaina”. In most cases, offsets were the same in the different arenas, so this helped to make it easy to reuse constraint definitions by simply changing the text attributes.

Only after these steps were completed was the event manager script file constructed to coordinate all the phase-related events. This information was entered into a single file, ready for testing. Even though at this stage we did not have all the final 3D geometries from the design artist, we were able to use cruder draft objects which were gradually replaced by the completed final versions as they became available, which involved only a number of minor changes to the world scripts.

3.2 Camera Controls

The previous section has described how appropriate world design plus constrained and managed participant movement were used to enhance the manageability of *Out Of This World*. In addition we created a specialised real-time camera control interface for the four camera operators. The characteristics of this interface reflect the real-time performance-oriented control requirements which derive from a live broadcast. Specifically:

- the camera can be used in a target based mode which allows the operator to keep the target in shot while changing the viewing parameters (e.g. to avoid occlusion and shots of empty space);
- camera targets can be static, based on hints in the virtual world definition, so that a camera operator can always jump to a known reference point if they are becoming disorientated, or if the activity is moving to a new arena;
- camera targets can also be locked on to individual participants or teams, ensuring that the cameras can keep pace with the action;
- the camera can also be used in a flying vehicle style, for maximum flexibility and control when required;
- variable damping is available to smooth movement and transitions;
- compound manipulations to the camera (affecting many parameters) can be specified which would not be possible with a normal widget-based or direct manipulation interface;
- adjustments to the camera can be pre-programmed and stored for later recall, or set up just before they are needed, and triggered as a single unit; and
- these compound actions can be assembled into longer “strings” of viewpoint transitions, for example to create extended camera set-pieces.

In the remainder of this section we will consider the construction and operation of this interface in more detail.

Figure 3.5 shows a screen image from one of the virtual cameras, showing both the normal flying vehicle-style controls (the arrow buttons at the bottom of the screen) and the custom camera interface (the dialog box at the right of the screen). The 3D window is sent to the mixing desk using the system’s normal video output capabilities (SG Octane, Impact and O2 machines were used for cameras). The 3D view is taken from the “space frogs” game: robot number one, in the background, has just chased a space frog towards the robot team leader, ready for it to be impaled on the team leader’s spiky hat!

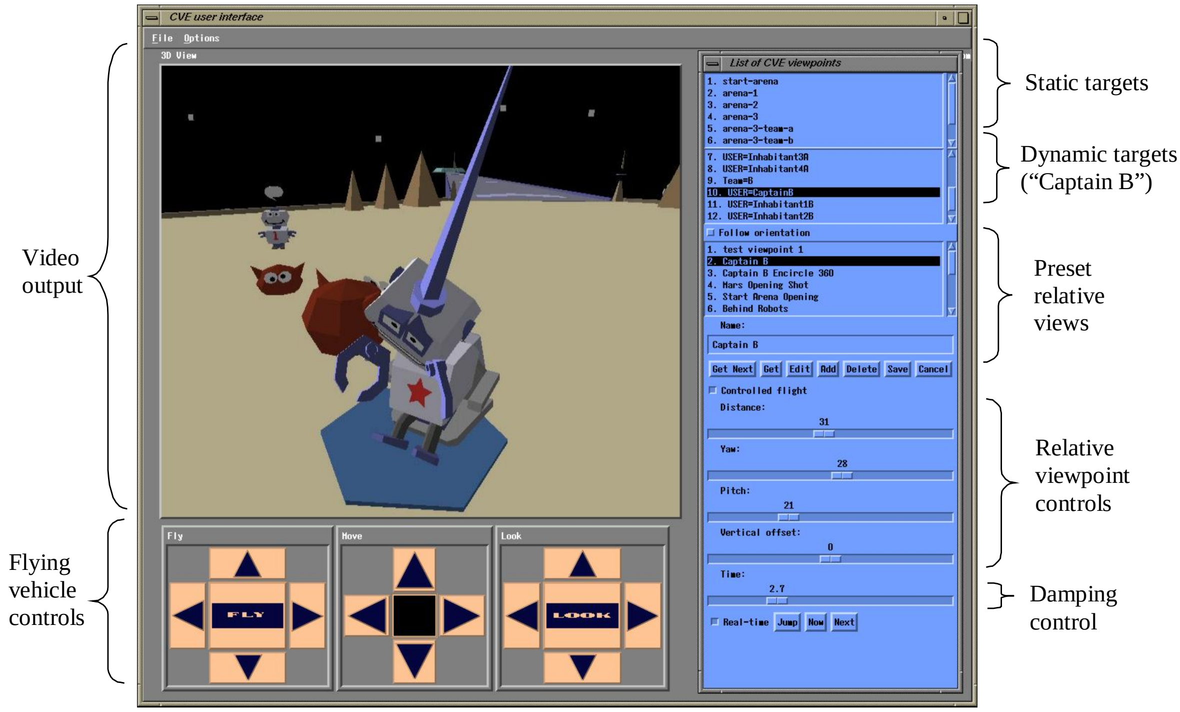

Figure 3.5: camera interface for Out Of This World

3.2.1 Flying Vehicle Control

The basic flying vehicle controls allow the camera to be moved in five degrees of freedom; the horizon was constrained to be horizontal at all times due to previous problems users had found with reorienting the viewpoint when rolling around the look vector. The main method of use is to control the camera's forwards/backwards velocity and left/right pan using the central black drag-box (which uses a cubic mapping from mouse displacement to vehicle velocity). Ware and Osborne (1990) found this navigational style (in their case controlled with 3D tracker) to be good for "movie making", with its smooth control and perceived flexibility. The flying vehicle controls can also be used in concert with the object-centered controls described in the next section.

As noted above, using only this interface in *Heaven & Hell – Live* produced a number of problems, such as keeping up with the action, getting lost and coping with occlusion in the scene.

3.2.2 Object-Centred Control

The flying vehicle metaphor controls the position and orientation of the camera. The alternative object-centred controls allow the camera operator to specify independently the target of the camera (the look-at point as in Blinn, 1988), and the relative viewing attitude and distance of the camera. This interface also provides additional temporal controls, as described below.

Look-at point control

The look-at point can be controlled in three ways:

- Vehicle-based free navigation, as described above, except that it is the target point rather than the camera itself which is being moved.
- Jump to preset fixed points in the virtual environment (derived from reference point artefacts included in the world specification), for example, game arenas and player starting positions. These are selected from the uppermost scrolled list in the right-hand interface.
- Continuously track a single participant or a group of participants (e.g. team A or team B). These are pre-configured options which are selected from the second scrolled list in the right-hand interface. The camera process continually monitors the activities of objects within the virtual world and adjusts the target as they move. In the case of group targets it takes the average position of the group's members.

It is worth observing that the first option offers direct control over positioning in 3D space, whereas the last two options are semantic choices expressed in terms of named objects or regions within the space. The inclusion of semantic choices allows the camera operator to jump to key locations or track participants or activities irrespective of their speed (both of which were problems in previous experiments). In this respect the system introduces elements of constrained and automated camera control (as in Drucker et al., 1992; and Seligmann and Feiner, 1991), except that it is the context of real-time control for performance.

In addition to specifying the look-at point, the selected target can optionally determine the reference orientation for the camera in the XZ plane (the ground plane). For example, when used with flying vehicle control over the target this allows the orientation of the camera to be controlled directly by turning the target. Alternatively, when used with the tracking option it allows the camera to adopt first-person perspectives (as the participant sees the world), locked facing shots (keeping the participant full-face at all times) and a number of other possibilities.

Relative viewing control

In the object-centred control mode the position of the camera is not specified directly. Instead, it is specified relative to the position (and optionally the orientation) of the camera's current target. The relative position is specified primarily in terms of spherical polar coordinates, an approach which we first made use of in *NOWninety6.* Specifically, the camera operator uses independent sliders to specify:

- the yaw of the camera relative to the reference orientation;
- the elevation of the camera;
- the distance of the camera from the target (including the option of going beyond the target to show an out of the eyes view for a participant); and

- a vertical offset relative to the target, to allow the operator to create views over or under the reference point, or to compensate for movements away from the ground plane.

These degrees of control are illustrated in Figure 3.6.

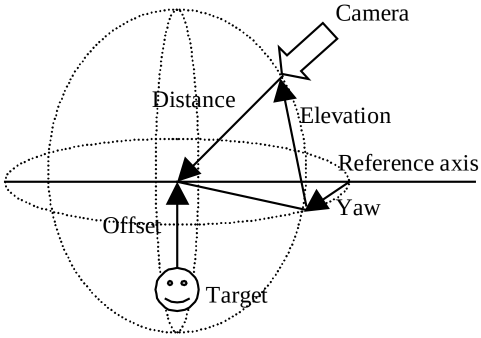

Figure 3.6: relative viewing controls

We suggest that this simple slider-based control over rotation is appropriate because it has been observed to be effective for single axis manipulation (Chen et al., 1988), and in our system we cannot use controls which overlap the 3D view because we cannot allow the mouse cursor to enter the graphical view (since it would be seen on the final broadcast).

Linear control mappings have been used for yaw and elevation. A logarithmic control mapping has been used for viewing distance, allowing controlled transitions between close and distant shots as proposed by Mackinlay et al. (1990). Vertical offset combines a linear central region (encompassing the expected range of ‘normal’ adjustments), with a logarithmic region at the extremes (allowing larger adjustments in ‘exceptional’ circumstances).

These controls together allow the camera operator significant (but not complete) control over the framing of the target within the shot. They also (we hope) make it relatively difficult to create incomprehensible or empty shots.

Temporal control

The final significant aspect of this control interface is its support for various forms of temporal control. The target point is controlled directly and interactively. However the relative viewing parameters can be controlled in three different ways, described below.

- Real-time: the camera moves as each slider is moved, subject to a controllable damping coefficient which allows the operator to trade responsiveness against smoothness of movement. The damping time constant is specified by the lowest slider control in the right-hand interface.
- ‘Just-in-time’: the camera operator can temporarily ‘disconnect’ the sliders from the camera, make coordinated adjustments to all aspects of the relative view control and then trigger those changes as a single atomic operation.

Again, the damping control determines the rate at which the camera interpolates between old and new settings.

- Pre-programmed: the camera operator can define complete sets of viewpoint parameters and store them for later recall. This is the function of the central portion of the interface, including the third scrolled list. These preset viewpoints can then be selected by the operator as a single action, triggering an interpolation to that new viewpoint (with the rate determined by the damping value in the viewpoint being recalled).

In addition, the just-in-time and pre-programmed modes also allow the operator to build up a sequence of camera movements, to be executed one after another. For example, the operator can select a number of camera transitions in rapid succession and have them executed sequentially, one after another (using the ‘Next’ button at the bottom of the interface). Alternatively they can discard any pending operations and immediately start a new interpolation (the “Now” button) or can jump to the new viewpoint with no interpolation (the “Jump” button).

It is apparent that temporal control over viewpoint in live performance settings can be somewhat more complicated than in non-real-time or non-performance applications. For example, in both computer animation and interactive exploration compound camera movements can be made incrementally (e.g. using a version of the real-time interface); it does not matter if the camera passes through intermediate ‘incomplete’ states. However in a live performance, every placement of the camera must be acceptable in itself, since it may be part of the broadcast.

The just-in-time interface allows the camera operator to compose complex camera movements to be executed as a single action. The pre-programmed interface allows the camera operator to reduce the amount of work which they have to do during the live broadcast by exploiting standard (or specialised) viewpoints which they have set up in preparation or rehearsal. Again, this is a direct result of the live performance nature of the interface’s use.

3.3 Initial Evaluation and Observations

This section details the immediate results gained from observing the performance, and comments from the world manager and camera operator personnel.

3.3.1 Event Manager

The management interface was very effective in imposing a staged and spatial structure on the event. In particular it allowed periods of essential autonomous activity to be interspersed with updates, briefings and transitions to new areas of the world and new activities. The management interface also allowed the local complexity of the world to be tailored to the activities at hand, for example the fish effectively disappeared from the system except for the time they were needed for the second ‘falling fish’ game. Other objects were controlled similarly. One technical problem involved conflict at times between the event manager and behavioural objects which sometimes refused to be reset. As a more general point this stresses the importance of correctly balancing autonomous and directed behaviours.

The audio nimbus interface allowed the audio 'power' of the different characters to the tailored before and during the event. By default, at the start of each performance the host was given a very large audio nimbus (5.0), the performers somewhat smaller audio nimbi (1.5) and the performers still smaller (0.7). At particular stages in the performance, or in response to events as they unfolded, the world manager dynamically readjusted these levels. For example, in focused dialogue the inhabitant concerned would be boosted, and when background audio increased the host and performers audio could also be boosted. However, Chapter 6 presents an argument that audio control may sometimes need to be even more dynamic than this.

Use of hierarchical spatial design helped maintain the world much more easily than in our previous experience, but it was noticed that:

- Changes often meant changes to more that one file.
- Lack of checking of uniqueness of naming text attributes could lead to design errors (removed by play testing).
- Lack of hierarchy/structure in event manager script file itself made it difficult to read.
- Little lower level control over infrastructure resources (e.g. Quality of Service) means it would not be easy to make changes if network performance changes. This was not really a problem in *OOTW* because a private Local Area Network was used.

The show itself was very coherent though this was undoubtably helped by the game-show content. However in one case the control was too severe - in one of the games, the users were constrained such than they could not have a "group hug" when it was requested by one of the team captains. This might suggest a need for more autonomous phases during a show.

As a final point, the event manager script needed to be constructed in coordination with the narrative script read by the game-show host, in order for the host to provide appropriate cues to the world manager (Chapter 6 makes further remarks about the relationship between these two scripts). This suggests a need to integrate script writing with both the storyboarding and event management parts of the design process.

3.3.2 Camera Control

The camera control interface was significantly more effective than the normal body-centred navigation which was used in *Heaven and Hell – Live*. In particular, the ability to follow individuals and groups allowed the cameras to keep up with the action in a way which they had frequently failed to do in that program. The object-centred navigation style also ensured that targets were consistently in shot, and allowed a range of static and swept overviews which supported the viewers' spatial understanding of the scene.

The provision of both the standard and enhanced navigation methods in the same interface was found to be useful as they were individually better suited for different tasks resulting in very different kinds of shots, and could also be used together.

Focus based work was found to be most useful when the operator knew what would happen. Usually this was only when the players movements were being controlled by the event manager—moving along the travellator, or stationary at the start and end of games. At these times the focusing was useful as it allowed interesting panning and pitching around a subject that could not be achieved with free movement (where you have to move the camera and look in the same direction). During the close-up interviews with individual inhabitants the focusing was especially useful for panning and pitching around the avatar, zooming in and out and moving up and down. Focusing on the inhabitant made up for their lack of movement and expression while they were talking. The focusing was also useful here for switching focus instantly when the next inhabitant was being talked to while retaining the viewing parameters e.g. a close-up on 1 became a close up on 2 instantly. This facilitated the tracking of conversation well. Switching between focus points at other times however, could not be done while that camera output was being broadcast because a view which looked good when focused on one inhabitant could look barren or cluttered when applied to a different target during a game. Switching focus during a game also looked confusing as there was often no relationship between the viewpoints.

The issue of focus switching applied to the switching between cameras as well. Initially, camera views were changed as soon as another camera had something more interesting, or the present view became boring. In later performances the operators learned to switch views far less often, allowing viewers to come to terms with what they were watching (for more details, see Chapter 6). Cameras were also switched while observing the same subject to provide a link and context. One camera would sweep towards a team or individual and then the view would switch to a camera focusing on it which would then pan and pitch around the subject—this allowed two cameras to co-operate to overcome the problem of switching from free movement to focus based work.

Using the free movement interface it was possible to view objects not listed as focus points such as the fish and space frogs which provided background to the game. It was also easier to pan from one team to another to provide relationship shots or just to follow the action. In these cases it was found to be less risky to use free movement rather than tying the camera to a single inhabitant, who may stop moving, move away from the action or do something completely unexpected. With the free movement control the operator could swing the camera around to find something more interesting rather than fumbling to find a new focus. Free movement also provided a greater variety of possible shots. Shots of a group of frogs with the players running towards them would have been more difficult using focus points. In the frog and football games it often looked better to follow a frog or the ball as a number of players fought over it rather than following a single player's efforts. These uses of free movement are discussed again in Chapter 6.

Using free movement in conjunction with the focus parameters was also found to be possible and often useful. In particular increasing the time parameter when using free movement made the movements smoother and eliminated camera shake caused by directly controlling the view with the mouse. The vertical offset and pitch control were also useful for setting a camera height and angle before a game which was then just moved horizontally. For example a low camera in the frogs game created more excitement and allowed the frogs to leap over the camera, while a higher elevation was needed in the fish game to see the fish and in the race as the car geometry was mainly horizontal and so looked better from above.

Additional observations were:

- The camera interface was very powerful, providing views that would be impossible with conventional cameras and the ability to instantly move infinite distances around the world, but it pushes the user towards very staccato camerawork that proved confusing to watch.
- It took time to learn how to use the interface to create flowing movements that were both more appealing and less confusing to watch. If the interface is to be developed it should be made more balanced allowing the smooth movements to be achieved more easily without sacrificing its current power.
- In game 2 the fish were given solid bounds so that they had to be navigated around on the ground after falling. The solid bound caused navigation problems. This may need to be addressed in the future, either by making the camera operators aware of solid objects, or by allowing cameras to pass unhindered through any object.
- It was easy to miss a potential shot by getting too close.

Improvements suggested by the camera operators were:

- Overview of environment, maybe a plan view. This would allow camera crew to see where the action is and where inhabitants are in relation to each other. This would inform choices for focus switching and provide directions for relationship shots which start on one team and need to find the other.
- View of broadcast output. This would provide a further clue as to when your view was being broadcast and allow you to find views complementary to those being broadcast - making the camera work more cohesive.
- Smoother switching between focus based and free movement. Being able to move toward an object then switch to focusing on them to pan around would be invaluable. The camera would be moved freely with the mouse, then a right click on an object in the current view would select the object as a focus. The problem with this interface in the present system is that the mouse would appear on the video output, but in future systems where the camera views just

provide viewing parameters for a single dedicated video output machine this would not be a problem.

- Time delayed focus switching would provide a link between two viewpoints, lessening any confusion caused by switching, and also providing context.

Many of these observations are confirmed by ethnographic observation as documented in the field study of *OOTW* reported in Chapter 6. This study also highlights issues to do with the relationship of camera control to direction and argues that enhancements to camera control software need to be seen in the general context of the cooperative activity involving the camera operators *and* the director. We return to questions of the enhancement of camera control also in Chapter 7, where an indication of current work by partners responding to these observations is given.

Chapter Four: WobbleSpace II - Audience Participation in *Out Of This World*

Jason Morphett
Advanced Perception Unit, BT Laboratories, Martlesham Heath, UK

4.1 Introduction

This chapter describes predominantly the WobbleSpace component of *OOTW*. WobbleSpace is an ongoing programme of research drawn from Workpackage 6 of eRENA. It is primarily concerned with using image processing and machine vision techniques to extrapolate information from the real world and propagate interactions into the virtual environment. For *OOTW*, this entailed applying the research onto activities to engage the audience in various forms of participation at two key stages in the performance:

Pong (warm-up) - an implementation whereby one half of the audience played against the other in order to 'warm-up' the audience;

Voting - a process whereby the audience got to choose one member of the losing team from the game to save from the *OOTW* 'doomed planet'

Both applications in WobbleSpace II differed in terms of the information presented to the audience insofar as during the *pong* activity, the audience were able to see themselves interacting with the game. Whereas during the *voting*, they were merely presented with an aggregate view of the overall activity. The implementation of WobbleSpace for *OOTW* took its inspiration from a broader field than purely machine vision. Indeed, designing for a participatory experience required the developers to draw upon heuristics for interaction design normally associated with situations which have successfully created sustained participation over time. Previous research in the field of audience participation is mainly through reflective experiences (particularly in theatre). However, beyond theatre, audience participatory activities have, and are, found throughout the world. We highlight a number of these here as influential when designing a participatory experience, in particular when generalising the activity within an Inhabited TV setting. Bilson (1995) cites the following as general guiding principles for participatory design which we have to a greater or lesser degree adopted:

1. Include content which is familiar and/or meaningful;
2. Use 'call and response' social rituals;
3. Think of improvisation as a growth path;
4. Encourage personal and physical expressiveness;
5. Acknowledge the individual;

6. Mark the passage of time;
7. Allow for a variety of roles.

Although not wholly applicable to the design of WobbleSpace, these principles are indeed a good set of heuristics for interactive and participatory design and we further cite other applications of participatory activity where these can be applied as well as returning to them again when describing the *OOTW* implementation of WobbleSpace in Section 4.4.

Current activity—post *OOTW*—has been informed through our post-event observations and evaluation. These have been presented previously to the project consortium at eRENA plenary meetings, and this chapter is a formal record of those presentations and results. Additionally, Deliverable D6.2 will report on subsequent WobbleSpace development, post *OOTW*.

4.1.1 Structure of Chapter

This chapter is organised into the following sections. Section 4.2 describes the influential factors in the design of WobbleSpace. In particular, version II that was used at *OOTW*. Additionally, it was our intention to evaluate the experience from both qualitative and quantitative perspectives. To that end, the evaluation section contains both personal reflections as well as invaluable feedback provided from the post performance discussion sessions and a presentation of the analyses taken from log files. We suggest improvements to the experience for both future developers of such technologies as well as recommendations intended to benefit the participatory 'experience' as a whole.

Section 4.3 describes the evolution of WobbleSpace as a whole as an activity under Workpackage 6 (i.e. 'Interaction') of eRENA. Indeed, *OOTW* gave eRENA an opportunity to practically realise underpinning research in a public demonstration—a key process in the eRENA project framework. Section 4.4 describes in depth the *OOTW* implementation of WobbleSpace, broken down by application into sections devoted to the design and development of both the *pong* and *voting* applications. We also describe our pre-production, set-up and calibration experiences and cite guidelines derived in retrospect intended to benefit future design and development. Section 4.5 is devoted to the evaluation of WobbleSpace. We summarise feedback from the audience discussions as well as our own reflections. Additionally, objective descriptive statistical analyses on the log files produced for both the *pong* and *voting* applications are presented. Finally, Section 4.6 sets out our conclusions and recommendations.

4.1.3 Acknowledgements

The research would not have been possible without the contribution of many parties. In particular, we wish to acknowledge Mark Jessop for technical development,

Jerry Bowskill for his invaluable input and the staff at the Green Room in Manchester for their patience and support in installing the equipment for *OOTW*.

4.2 Audience Participation

Bilson states that interaction design, in the realm of electronic participatory experiences, can draw inspiration from real world live performances within cultural and social contexts (Bilson, 1995). In his article ‘Get into the groove: designing for participation’, Bilson describes the relationship between live performance and audience participation, focusing particularly on examples in which the performance is catalytic for *community building*. He cites examples from theatre, music, parades and demonstrations, religion and schools as examples of real world participatory activities. We took Bilson’s guidelines as set out in his paper as instrumental when designing WobbleSpace II for *OOTW*. Each of the guidelines is more fully described in Section 4.4 where we relate them to WobbleSpace II’s design. This section however, more broadly elaborates on audience participation *per se* and upon the role of audience participation in production and performance, also citing other examples of its use.

4.2.1 Background

A survey of the literature on current implementations of audience participation systems reveals applications ranging from music [HREF1], theatre [HREF2], opera [HREF3] to devices aiding meeting planning [HREF4] and theatre ‘warm-up’ activities [HREF5]. Of particular interest to us in designing WobleSpace II is the CINEMATRIX system [HREF5]. This system incorporated a combination of laser and image processing technology allowing large theatre audiences to interact with one another using a wand (or paddle).

This system achieved much acclaim when it was first shown at SIGGRAPH in Las Vegas in 1991 and more recently at Ars Electronica (1998) at Linz [HREF6] and finally, again at SIGGRAPH in 1998. It also has permanent installations at Parc du Futuroscope, Poitiers, France and LAVA of Hamburg, Germany. The system allows audience members to interact with the screen to play games, answer questions, create patterns and make decisions about adventures in a virtual world, usually in competition by dividing the audience into components. Of interest to our system was the fact that the interaction does not require electronics nor complicated devices to interact. Indeed, the wand device needs no button pushing or wires to operate it. Moreover, the device is merely held up by the audience member with the appropriate colour facing a laser for the interaction to take place. We sought to extend this style of interaction through asking the audience to *wave* the card (used instead of a wand) thereby causing the activity to become more animated than just holding up the ‘prop’—see Section 4. It was envisaged that this would encourage the audience to become more active and thus enjoy the experience more fully, which it seemed to achieve when we measured the level of activity and interest of the audience—see Section 4.5 and Figure 4.8.

The following section more fully describes the issue of *methods of interaction* for audience participation. In particular, we extend the description of interaction for WobbleSpace and justify the decisions taken when considering participatory interaction in a theatre setting.

4.2.2 Methods of Interaction—Group or Individual?

Participatory interaction requires a ritual to be in place so that both the audience know what to do and the receivers know when to act upon it (Bilson, 1995). Exceptions to this rule are improvisory theatre whereby the audience (potentially) directs the production or performance in a non-linear fashion. However, in one example, the Rocky Horror Picture Show (RHPS) shown originally at the Waverly theatre in New York in the late 1970's, this impromptu method of interaction grew into a collective ritual that the regular RHPS community later adopted as part of the performance. Indeed it could be argued that without this participatory activity, the RHPS may not have evolved into the institution that it is today. Initially, the RHPS used a process of audience participation merely to 'warm-up' the audience and encouraged the theatre members to join in. This they did to the extent that over a period of time, a quirky dialogue between the audience developed into a culture that allowed, and implied, that the audience's interaction over time, became part of the legacy that is now revered as one of the major exemplars of a successful participatory production. Piro [HREF2], described the RHPS community activity and the phenomenon from it roots in 1977. Interestingly, this revealed an emergent cult community born from an affordance—and an eventual encouragement—for the audience to participate. Of particular note, was the strength of bond both felt and fostered by the community.

The methods of activity ranged from initially shouting quirky responses at key moments, to collective singing and chanting and in the later years to 'dressing up' and prop handling. What is evident from this is that a community enjoyed the experience of the production because they were empowered with (and eventually expected) to be a part of the show. However, they could not influence it—something WobbleSpace II could.

Other societal real world examples of methods for collective interaction include dancing, such as cited by Bilson during South American Rumbas (Bilson, 1995) and the following of characters in a so called 'multiple point of view dramas'[1].

What is evident from all these examples, is that the experience of the individual *may* be dependent on the level of the collective experience and on the amount of impact that

[1] Bilson cites 'Tina's Wedding' as an example of this form of theatre.

activity has on the performance. Indeed, the experience (or level of interest) appears to stem from one or more of the following perspectives:

1. the audience as a collective is being involved;
2. an individual from the audience is being involved;
3. the method of interaction uses a device;
4. the interaction is impromptu or ritual based;
5. the interaction is form based (dance or movement) or auditory (sung or chanted).

Of note, is the potential for these to become combined together interchangeably. For example, interaction being impromptu and an individual from the audience being involved - as in the case of improvisory comedy. Or, consider that the participation could be a ritual and the audience as a whole is involved—as in the case of the RHPS.

However, for WobbleSpace, we sought to experiment with the combination of an audience as a collective being involved (1. above) using a device (3. above).

4.2.3 Summary

This section has briefly introduced the issue of audience participation citing guidelines for design as well as describing some of the more recent activities involving the public in performance. With this in mind, the next section briefly departs from audience participation to give some technical background to the WobbleSpace development as a research activity. We will return to the issue of participation, and in particular design for participation in Section 4.4.

4.3 Evolution of WobbleSpace

WobbleSpace is the vehicle we have used to explore the use and role for video in collaborative virtual environments beyond that used for conferencing. In particular, we have sought to apply image processing techniques in order to reflect, through either inference or explicitly what the camera 'sees'.

4.3.1 Image Processing and Machine Vision

For centuries, man has been interested in solving the puzzle of how we come to 'see'. The first computational experiments in developing artificial machine vision systems were conducted in the late 1950s and, over the last 25–30 years computer based vision systems of widely varying degrees of complexity have been used in areas as diverse as office automation, medicine, remote sensing by satellite in both the industrial and military world.

Machine vision is a multidisciplinary subject, utilising techniques drawn from optics, electronics, mechanical engineering, computer science, and artificial intelligence.

Machine vision is (in general) broken down into three distinct phases: image acquisition, image processing and image analysis.

However, during the WobbleSpace experiments we have been primarily concerned with processing the image data and interpreting results coarsely rather than applying the artificial intelligence algorithms to *interpret* the data. Thus, rather than discussing and developing an application that deals with the complex interplay between visual perception and motion control—i.e. action—we have chosen to focus on a more abstract—and computationally efficient use for machine vision in the context of linking the ‘real with the virtual’—particularly under Workpackage 6.

The following sections focus upon the major stages in machine vision mentioned above, briefly summarising the components required. We will return to these headings later in his section as a means by which to describe each of the WobbleSpace developments, particularly focusing on the *OOTW* application later in Section 4.4

Image acquisition

However, before we discuss image acquisition, a brief summary of *illumination* is required. Particularly given our experiences at the Green Room in set-up and calibration of WobbleSpace II (i.e. the *pong* and *voting* applications at *OOTW*). The particulars of which are further discussed in Section 4.4

Both *scene* and *object* illumination play a key role in the machine vision process. the central purpose of imposing controlled constant illumination is to visually enhance the parts to be imaged so that their flaws, defects and features are highlighted (Vernon, 1991). Incandescent light bulbs (such as those used at *OOTW*) are generally a cost effective source of illumination. However, they are generally undesirable for image based systems due to their directional lighting characteristics. Hence, incandescent lights tend to cast shadows over the image that ultimately interferes with results. Although the Green Room in Manchester did indeed support diffused bulbs, the image capture process was affected by shadows. Additionally, incandescent bulbs emit infra-red radiation and although not harmful to humans, again interferes with CCD (Charge Coupled Device) cameras—such as those used at *OOTW*. Finally, controlling the amount of light into the camera at *OOTW* was also important.

As many inspections systems—including WobbleSpace—base much of their analysis on the absolute intensity—or luminance—of the incident light, the control of the sample area illumination is important. Light meters or self-calibrating mechanisms can be used to aid this problem and are indeed used in manufacturing. However, in *OOTW* we decided to manually calibrate each camera such that the aperture and capture frequencies for both cameras were identical to minimise the noise differential between them. Indeed, both cameras were auto-iris disabled and we manually overrode the settings such that both cameras received the same amount of light as we were using a multiple camera configuration.

Having covered the problems of illumination, we turn our attentions now towards the process of image acquisition. The analogue captured scene is typically converted to a digital representation through a process known as *sampling* and *quantisation*. The image is held as a two dimensional array of pixels (represented as integers in our case as we chose to use 16 bit colour). These represent the reflectance function of the actual scene at discrete intervals (i.e. the *sampling* rate). The difference between these 2D arrays is used in image analysis later—post processing—and determines the amount of spatio-temporal difference between samples. The process of *quantisation* is concerned with converting the analogue signal captured by the frame grabber into the 2D array of discrete values as pixels

Image processing

Although the distinction between image processing and image analysis is not immediately obvious, it is an extremely important one. Image processing can be thought of as 'a transformation which takes an image into an image' (Vernon, 1991). On the other hand the process of image analysis is a transformation of an image into something other than an image, i.e. it produces information representing a description of the image.

Broadly speaking, there are three typical classes of image processing:

Point operations;

Neighbourhood operations;

Geometric operations.

As WobbleSpace used only *point* and *neighbourhood* operations, we refer the reader to Pitas (1993) and Vernon (1991) for a description of geometric operations.

Point operations are concerned with operating on each pixel in the output function using (normally) grey scale values. In the case of WobbleSpace though, we used RGB (Red, Green and Blue) and HLS (Hue, Luminance and Saturation) 16-bit values as discussed further in Sections 4.3.2 and 4.4. However, in both cases, *thresholding* was a key point operation used.

Thresholding is used to extract pixels that fall either above or below a certain value in the array. In our case, we used thresholding to seek out (i.e. *segment*) pixels that met certain RGB or HLS values ± an offset. Additionally, for *OOTW* and WobbleSpace II a neighbourhood operation was employed.

In the main, neighbourhood operations generate an output pixel on the basis of the pixel at the corresponding position in the input image *and on the basis of its neighbouring pixels*. Typically, the size of the neighbourhood array is 3×3, but some applications may extend this to 5×5 or more. For WobbleSpace II though, we chose a 3×3 neighbourhood array as this was sufficient for our needs.

Neighbourhood operations are also sometimes referred to as *filtering* or *noise reduction* operations. WobbleSpace II employed a *median filter* in its image processing algorithm to achieve this. This technique assigns the output pixel as a function of the input neighbourhood by sorting the array values and computing a median output value. The effect of which tidies up spurious 'noisy' pixels from the image. It is considered a superior technique to *mean filtering* as it reduces the effect of blur—an advantage we subsequently lost due to a reduced capture resolution (¼" PAL) at *OOTW*. Other techniques for noise reduction exist, such as Gaussian filtering, Convolution, Thinning, Erosion and Dilation[2]. However, WobbleSpace didn't use these as the median filter alone was sufficient.

Image analysis

As mentioned, *image analysis* is concerned with the extraction of explicit information regarding the contents of the image—i.e. we are seeking to *reason about the data*. Both implementations of WobbleSpace to date have tended to provide explicit rather than implicit representations gleaned from the processing the image pixel arrays—i.e. we are not inferring an interpretation, moreover we have represented that which is known in an abstract way.

In particular, luminance has been a key component in the determining the final representation. More 'intelligent' uses of image analysis include Template Matching, Blob and Optical Flow Analysis and Histogram Analysis[2]. However, in WobbleSpace, the analysis has been a more fluid (and cost effective) operation of interpreting spatio-temporal differences between samples. Additionally, WobbleSpace has been configured to operate over two components across the network. The server is the main 'image engine' and the client a receiver of aggregate information pushed by the server. The server performs the image capture, processing, analysis and packetisation of data which it then publishes to all connected clients. In the case of *OOTW*, MASSIVE-2 (Greenhalgh et al., 1997) was the client and a process of cross platform integration was required to allow WobbleSpace to message MASSIVE-2.

[2] For a fuller description of these techniques, see [Vernon91] and [Pitas93]

The following sections more completely describe the client-server process through the evolution of versions of WobbleSpace demonstrated at the i3 conference at Nyborg and the *OOTW* implementation of WobbleSpace.

4.3.2 WobbleSpace I (Nyborg)

The original function of WobbleSpace can be defined as, "To investigate the effect of aggregating video crowd activity and its perceived synthetic and auditory representation in a Collaborative Virtual Environment (CVE)"[3]. This is the remit as defined under the auspices of workpackage 6 for the exploration of *linking the real and virtual* in a non-intrusive way. In particular, we sought to combine video with virtual environments in new and novel ways, beyond that used for conferencing purposes. This section briefly describes the early developments on WobbleSpace I with the following section more fully describing WobbleSpace II as developed for *OOTW*. However, we feel it necessary to provide some background research before departing onto more current development.

As alluded to, previous research into the use of video inside virtual environments has predominantly focused on the supposition that video is used for forms of inter-inhabitant communication, i.e. video conferencing. Indeed, our previous work on eTV (Morphett, 1998) more fully describes our own experiences of using video and virtual environments for this purpose in what we term 'spatial video conferencing', other examples of this type of use can also be found in Reynard et al. (1998), Frécon et al. (1992), Han and Smith (1996) and Suzuki (1995). The reason for this use of video and virtual reality has primarily been due to research into video conferencing and media spaces (Bly et al., 1993). However, there are potentially more abstract awareness driven uses for video in CVE's. Therefore, WobbleSpace I demonstrated how 3D animated objects and sounds, could be used to *infer* video activity; e.g. such as a crowd cheering enthusiastically. And, to a lesser extent, what the virtual adjunct would look/sound like.

By way of analogy, consider the graphic equaliser found in most hi-fi equipment which shows an audio signal amplitude variance, even if the volume is turned down. By turning up the volume on the amplifier, the auditory source is heard. WobbleSpace I provided both visual and auditory cues that could be used inside a CVE, such that users attention and awareness could become attracted to these 'agitated' objects. Additionally, the source interaction is non-invasive and passive, i.e. activity was registered by just moving in the cameras field of view. Also, the final representation was abstract seeking merely to *attract attention* rather than *convey information* about what was happening.

Indeed, it can be said that WobbleSpace I was concerned with representing the fact that something *had* happened rather than *what* had happened. And, although WobbleSpace

[3] This is extracted from the description provided for the Nyborg i3 conference introducing WobbleSpace.

was still very much “work in progress”, it was hoped that it would be able to provide answers the following questions:

- Could we (automatically) extrapolate behaviour from video?
- Could we abstract these behaviours such that they are still perceivable in the VE?
- Could we apply a spatial model to the virtual adjunct to moderate awareness of the object?

The WobbleSpace I implementation was divided between two software components: A 2D component (the server) and a 3D component (the client). The 2D server captured video images from a camera (this happened at ~15Hz) and applied some rudimentary image processing to provide an aggregate of the captured images—i.e. luminance differencing based on macro-blocks.

This implementation divided the captured images into four equal quadrants and used macro-block luminance differencing to calculate the number of pixels that had changed in each quadrant. These values were then published on a well-known TCP/IP address.

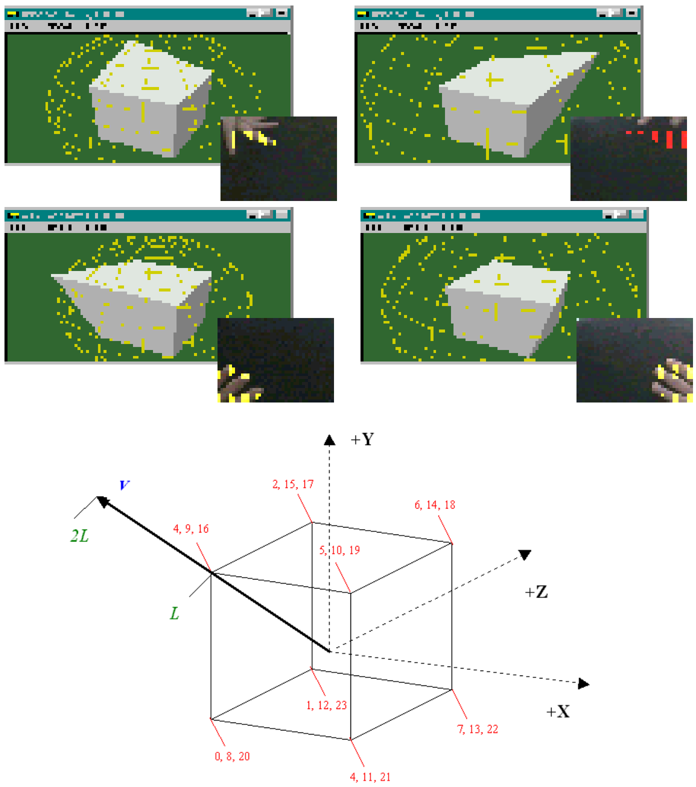

Figure 4.1: WobbleSpace I showing the cube morphing process.

The 3D client represented the video stream with a cube and an encompassing sphere to represent the aura (Benford and Fahlén, 1993). A single vertex was mapped to each of the four aggregate values that were provided by the server (see Figure 4.1 above[4]). Each of these vertices also had an audio sample associated with it. When aggregate values were received, the client used the values to morph the four corresponding vertices of the cube and change the frequency, gain and balance of the associated audio sample. The total amount of change over the entire video image, i.e. the sum of the aggregates, was then used to vary the size of the object's nimbus, i.e. the yellow sphere in Figure 4.1. Processing the incoming four values from the packet proceeded for each of the vertices we wished to modify. We chose to vary the vector lengths in Figure 4.1 between L and 2L for each top vertex of the cube, such that:

Assume you want final vector length Vf from the incoming video quadrant value Qv :

```
normalise V – giving Vn.
Where: V = the scaled vector in fig.1
Sc = (Qv+256.0) / 256.0
This gives Sc as a scale factor varying between 1-2 as 0 >= Qv <= 255
Then
     Vf = Vn * Sc * L.
     |Vf| is now be between L and 2L
```

This process not only provided a way of abstractly representing a moving video stream, but also a way of reducing the unnecessary use of bandwidth used in transmitting the actual video stream until objects awareness (Benford and Fahlén, 1993) was sufficient to trigger the stream. The full video stream would then only be transmitted once inhabitants fell within the objects aura and therefore volunteered some 'interest' in the content.. Due to the problem with using luminance difference as the *only* image processing technique, we were going to deepen the algorithm in order to make sense of what the camera saw. Figure 4.2 below illustrates this problem. Here, both +ve and –ve differencing produced the same result. Note the dialog's 'Avg. Difference' edit box values. However, WobbleSpace II was again only interested in representing that activity *had* happened and not *what* that activity was.

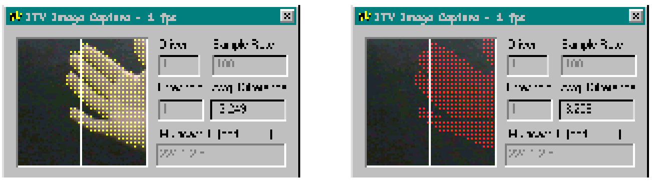

Figure 4.2: WobbleSpace I server showing the problem of using luminance differencing alone.

IMAGE AQUISITION: Winnov™ conference based PC camera

[4] Note vertices 4, 9, 16, 2, 15, 17, 6, 14, 18, 5, 10 & 19 are morphed. Consider Figure 4.1 and that video quadrant 1 morphs vertices 2, 15 & 17, quadrant 2 morphs 6, 14 & 18, quadrant 3 morphs vertices 5, 10 & 19 and quadrant 4 morphs vertices 4, 9 &16.

IMAGE PROCESSING: Luminance differencing, algorithms by application author

IMAGE ANALYSIS: Mapping quadrant differences to vertex morphs (as above)

4.3.3 Summary

This section has described the background to the WobbleSpace research activity. In particular, we have described WobbleSpace I's development and evolution. The following section returns to the issue of audience participation and more fully integrates the heuristics from Section 4.2 in the development of WobbleSpace II for *OOTW*.

4.4 WobbleSpace II (*OOTW*)

4.4.1 Design, Development and Testing

We took the basics from the WobbleSpace I demonstrator shown at the Nyborg i3 conference as the technical point of departure for the design of WobbleSpace II. To this, we applied Bilson's guidelines for participatory design (Bilson, 1995).

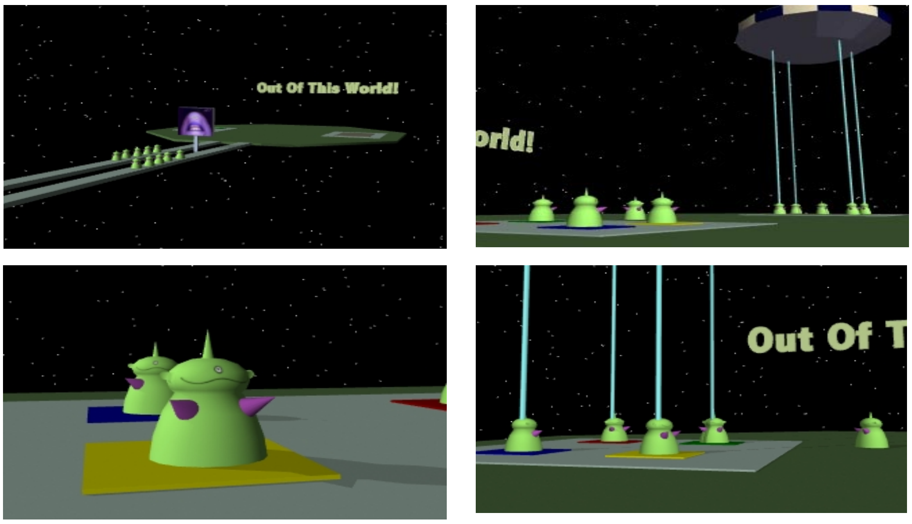

Figure 4.3: WobbleSpace II – early screen shots taken from the animated storyboard

Include content which is familiar and/or meaningful

Familiarity with the content of an experience makes it much easier to participate. In religious domains, if the subjects know the prayer, they are able to recite it; in demonstrations, if the congregation know the rallying cry, they can add voice to the throng. In the *pong* application in *OOTW*, we chose a game with historical meaning that was obvious to the audience. The rules of *pong* are simple, manipulate the bats such that you defend your own goal area whilst trying to score against the opponent. Indeed, *pong* has been a popular choice for other audience participatory 'games'. The best example being the CINEMATRIX SIGGRAPH experience [HREF5]. Here, audience members were asked to show either red or green paddles that either moved the *pong* bat up or down. A similar principle was employed by the WobbleSpace *pong* game. The difference being that during the *OOTW pong* game the audience were able to *observe* themselves through a mixed reality display projecting both themselves interacting as well as the game combined into the broadcast.

This was intended so that we could evaluate differences in interaction behaviour from the *voting* mechanism, when all that was shown was the aggregate effect of the audiences activity. We shall return to this point again in Section 4.5.

Use 'call and response' social rituals

Call and response is a pervasive and time tested social ritual that strongly facilitates participation. It refers to the dialogue between the 'group' leader and the followers, with a lead chant being accompanied with a group chant. For example, chanting at a peace or other such rally, e.g. 'what do we want?' with the response being 'freedom' etc. This rhythmic turn-taking repetitive activity places firm boundaries between cause and effect.

In *pong*, the desired effect was that after a period of experimentation with the cards, the audience would encourage each other to wave either red or green to move the bat up or down—as observed by a colleague from the SIGGRAPH experience. In effect the behaviour observed was quite different as covered later in Section 4.5.

Think of improvisation as a growth path

Comedy improvisation – another example of an audience participatory activity—uses improvisation, or content deviation, as a growth path. Audience members feel part of the experience through being able to steer the activity in a direction that they believe they have control over. Indeed, the growth path is an audience empowering concept, passing flow control over to the audience at intervals throughout the production.

For our part, during the WobbleSpace *voting* phase, the audience had most of the control over the speed at which the voting happened. During the rest of the performance, MASSIVE-2's production interface commanded *inhabitants* into start points and only moved them forward when the producer decided to allow it. During the *voting* however, this was not the case. The audience were encouraged to vote for their 'favourite' losing team member through waving a colour coded card. At that point they had relative control over the show's speed in terms of defining the rate at which the voting phase took. This is in contrast to traditional TV, whereby the director 'stage manages' the sequence of events according to a 'running order'. In summary, passing control to the audience at this point was intended to make them feel that they had some control over the content.

Encourage personal and physical expressiveness

Speaking is more expressive than being silent, animated behaviour is more expressive than mute activity. For example, in instances of child workgroup projects, children are observed moving around whilst talking with one another. In another example, during evangelical preaching, the congregation are frequently seen to augment the experience through waving their arms and standing. The design of the *voting* mechanism (and to a lesser degree, the *pong* activity) was specifically designed to encompass this behaviour. Perhaps more appropriate for a theatre setting than for an Inhabited TV event, but nevertheless, an important factor to consider when designing social participatory activities.

Rather than just holding up the card to register a vote or a score such as at SIGGRAPH, the participants were encouraged to wave 'vigorously' in order for their votes to be registered.

Indeed, it was our intention to allow, for example, a higher score for *X* number of people waving more 'vigorously' than *2X* people just holding up the card. Section 4.5 describes the effect of this in more detail. However, suffice to say that we experienced periods of people standing and waving vigorously in response to our request whereas otherwise we could have perhaps have not seen such animated activity if we were just registering 'static' votes—i.e. just holding up the card, or pressing a button.

Acknowledge the individual

In traditional interaction and interface design, feedback is an intrinsic element in engineering usable systems (Preece, 1994). This also applies to live performances and street theatre as well as television. Consider the game show on television where the audience are asked to vote for a winner. A measure of the current state of the voting keeps the audience in suspense until the winner is announced. It helps encourage the audience to vote for their favourite and in some cases allows for a parallel 'race' to occur between the winners and losers sometimes allowing the rate of *voting* to change over time, i.e. competition amongst *viewers*. This was another very interesting event for the *OOTW* experiment in terms of measuring activity rates over time given a situation whereby in the first instance the audience were able to see themselves and the other, where they were not. Additionally, during the *voting* activity, the audience were able to dynamically alter their decision in a way that a digital *voting* system wouldn't allow. This afforded a race potential that made for some interesting results—further discussed in Section 4.5.

Mark the passage of time

This is a more pronounced guideline in an ongoing activity. It is concerned with sustaining and presenting key moments in time with a ritual. For example, religious social activities are punctuated with congregations assembled for specific purposes, e.g. to mark the birth of Christ at Christmas etc. Regular television game shows and talent competitions usually mark the passage of time with a 'top scores' section from previous weeks etc. However, for *OOTW*, as it was in essence a one off event, there was no requirement to mark the passage of time. However, should the concept of Inhabited TV become commonplace in the future though, and in particular if it involves some participatory activity, it would be wise to adopt this temporal activity to stimulate the audience and inform them of past activities.

Allow for a variety of roles

This heuristic is aimed at accommodating multiple levels of experience within a group and is primarily aimed at games design. So again, WobbleSpace II wasn't concerned with this guideline.

In summary, WobbleSpace II adopted the following design goals. Please note that for the following, a trailing 'p' refers to the *pong* implementation and a trailing 'v' refers to the *voting* implementation:

Design the activity so that the gameplay is easily understood by the user [p]

Make the activity engaging by providing feedback to (i.e. acknowledge) the individual [p]

Make the effect of the activity obvious and simple [p&v]

Encourage activity and personal/social expression [p&v]

The two applications that constituted WobbleSpace II (i.e. *pong* and *voting*) were specifically designed to be subtlety different to enable us to gauge and reason about disparities in the observed behaviour. Each of these implementations is further covered in later in this section.

4.4.2 *OOTW*: Image Acquisition

Turning our attentions now to the physical implementation of the software and hardware for WobbleSpace II, we now discuss the technical implementation in support of the preceding design goals. Initially, WobbleSpace I used cost effective 'conference' cameras to capture images. The manual configuration of which was minimal as was the capture resolution and Auto Gain Control (AGC).

Of primary concern was an inability to disable the auto iris. This precluded us from sustaining a constant reception of light into the device, especially important when using a multiple device configuration. We also had limited control over other features such as capture rate, colour space format and in particular, no support in terms of an API. We therefore decided to purchase a CCD camera with the following specification

Image Pickup	½-inch, interline transfer CCD
Effective Pixels	44,000 pixels [752(H) x 582(V)]
Scan frequency	15.625kHz (H), 50.0 Hz (V)
Horizontal resolution	470 TV lines (H)
Minimum illumination	0.95 lx (25%, F1.2, AGC “18Db”)

Additionally, the WobbleSpace I image processing algorithms were coded manually doing all the point and neighbourhood operations in the application software. The purchase of a Matrox Meteor 2 frame grabber allowed us to utilise the services of the Matrox Imaging Library from Matrox. More specifically, The *MIL light*™ version allowed us enough functionality to accomplish the technical requirements. Thus, a complete rewrite was needed from WobbleSpace I to WobbleSpace II to accommodate the new API. However, substantial gains in terms of software reuse were permissible through the use of the ITV libraries developed as part of WP3 of eRENA. Particularly in terms of an accelerated path for the TCP/IP networking code.

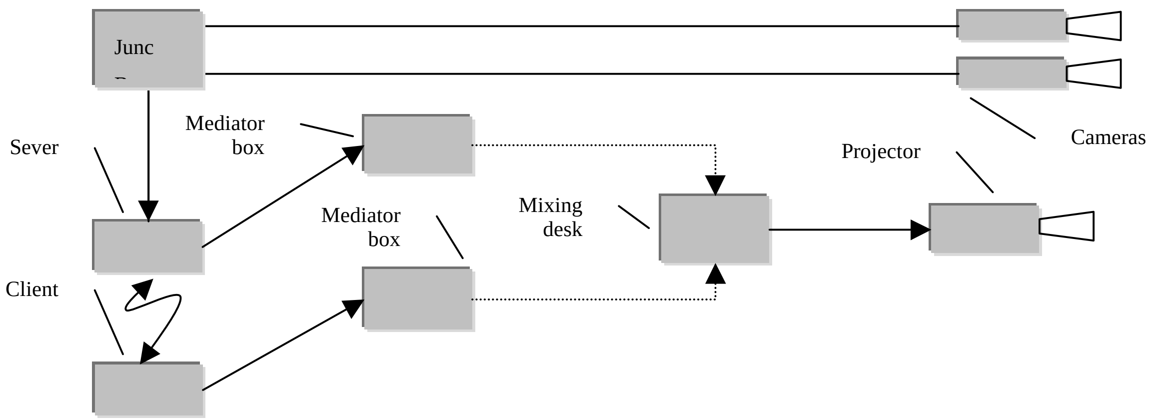

Figure 4.4: WobbleSpace II – configuration schematic.

Thus, our image acquisition hardware consisted of 2 JVC TKC1380 cameras equipped with 6mm lenses (following a feasibility pre-production visit to the Green Room by the developers). The cameras were configured identically to minimise noise and disparity at capture time, i.e. auto iris disabled, frequency synched and Auto Gain Control (AGC) disabled. The cameras were then connected to an s-video signal junction box where the ¼ PAL capture images were composited into a single PAL image for subsequent processing. The junction box was then connected via coaxial cable into the PC server via the Matrox Meteor 2 frame grabber and was programmed using a combination of C++, Win32, MFC, our own ITV libraries (Morphett, 1998) and the MIL library.

4.4.3 *OOTW*: Image Processing

As the WobbleSpace II *image engine* (i.e. the server) was in effect only communicating with a single composite image in the junction box, we didn't need to synchronise the cameras in software. Indeed, this had already been done in the cameras, so we could ensure that the capture rate of each camera into the junction box was identical. Therefore, all that was needed was to command the junction box, via the Meteor 2 board to quantize the image data and transfer the results into the Meteor for processing.

Once we had the 1st image buffer in memory for processing, the next phase was to segment the image such that we removed all pixels that we weren't interested in. This came through a process of a bitwise HLS ± offset operation on the 1st buffer with known HLS values in the search array. This search array was populated during the calibration session whereby the coloured card was held in front of the camera to pick out the 2 or 4 colours that we wished to search for (2 for *pong* and 4 for the *voting*). The search array held values, and the 1st image buffer was searched for corresponding matches to the input pattern. If a match was made, then for each of the target colours a binary masking buffer was populated with true values (i.e. found) or false values. Once each of these 2 or 4 masking buffers were set up, we filtered each of the masking buffers with a median filter.

At this stage we now had an initial image buffer (for each camera) and four masking buffers with identifiers in the 2D array showing matches to known HLS target values. Each of these masking buffers has also been subjected to a median filter to eradicate spurious pixel data.

4.4.4 *OOTW*: Image Analysis

The process of image analysis used the masking buffer information. It also had a copy of the last masking buffers that it used as a datum to infer the total mount of unique activity. To highlight what we are analysing, consider that holding a red card in front of the camera for successive frames would not register values. However, moving the card modified the current and reference frame buffer values (i.e. the masking buffers) and did indeed register values as there is spatio-temporal activity in the scene. The application of an XOR operation between the reference and current masking buffers yielded a pixel score for each colour that identifies the amount of new activity between successive frames. These values are further discussed below under both the *pong* and *voting* sections. Then, for *pong*, the masks were used in conjunction with specific HLS values for each of the colours and an operation that cycles through the 2D masking array looking for 'true' pixel fields. It then copied the contents of the relevant HLS pixel over the original image. The composite image was then piped from the PC into the TV production mixing desk (in the case of *pong*) for broadcast (see Figure 4.4 above). Finally, the contents of the current masking buffers were copied into the reference buffers and the cycle repeated.

Figure 4.5: WobbleSpace II – early lab prototyping showing the process of splitting the image in two for each side of the audience (for Pong) . During OOTW, two camera images were composited in a junction box. N.B. HLS pixels from the masking buffers are written into the capture buffer.

4.4.5 Pong Application

Originally, the *pong* section of *OOTW* was merely intended to act as a warm up for the audience as suggested by the director. However, it actually enabled us to experiment with alternative forms of representation in order to assess whether the audience reacted differenctly between *voting* and *pong*.

Historically, ‘pong’ was originally a two player arcade game. Nolan Bushnell, with Ted Dabney, invented the first video game. That game, however, was called not pong, but ‘Computer Space’, which Bushnell described as “a cosmic dogfight between a spaceship and a flying saucer.” Marketed in November 1971, it did not sell well. Later, Bushnell said it was too complex for 1971’s video game players.”

Undaunted, Bushnell and Dabney, as well as Al Alcorn, formed a company called Syzygy, and designed pong. They renamed the company Atari, and put pong on sale in November 1972. The arcade game pong became wildly successful. Atari Sold approximately 10,000 units, while at least 25 other companies sold about 90,000 copies or adaptations.

We chose to use *pong* as it is a simple game and the rules are well understood. However, the challenge then became one of how to allow the *OOTW* audience to play a two player game with only coloured cards. As mentioned previously, CINEMATRIX achieved this with lasers and coloured paddles [HREF5]. However, as mentioned, we intended to vary the interaction slightly to allow the audience to physically express themselves—i.e. through waving coloured cards vigorously.

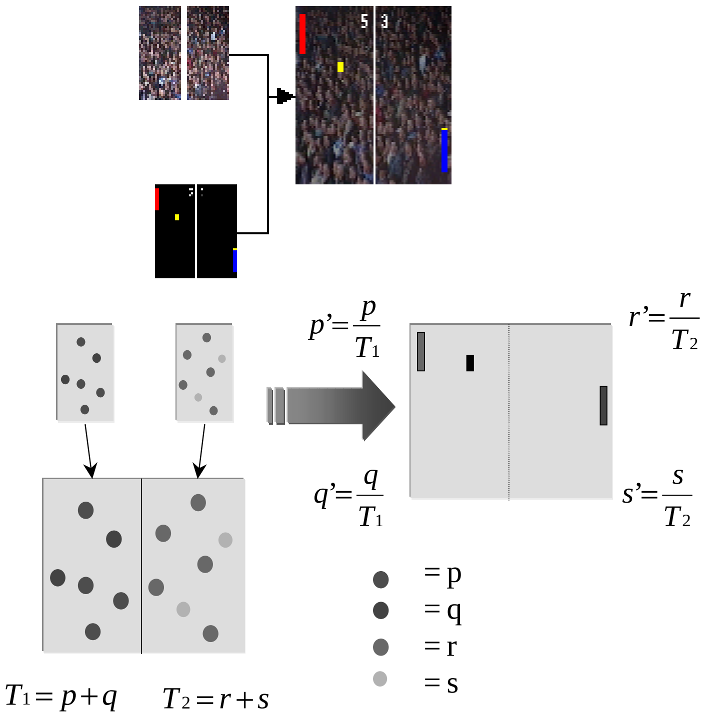

Figure 4.6: WobbleSpace II – Pong processing showing the views of the cameras combined with the game (top) and the processing done by the image engine, i.e. the server (bottom).

Therefore, we used the image engine (i.e. the server) to grab frames from the junction box at 10–15Hz. Each frame was then segmented and the four colours (lets call them *p*, *q*, *r* and *s*) were then extracted and the total number pixels were used in calculating *p'*, *q'*, *r'* and *s'* as above in Figure 4.6.

These prime values were then used to move the paddles a proportionate amount (call this *D*) according to the strength of the prime values such that (in pseudo-code) for the red bat:

```
IF (P' > Q') D' = D * P'-Q'
ELSE D' = D * Q'-P'
```

The red bat was then moved the proportionate amount (i.e. *D'*). Finally, to allow for varying audience sizes, a period of calibrating *D* at run time was necessary. This took in the order of 5–10 seconds to achieve so that the amount the paddle moved per frame was playable for varying audience sizes.

4.4.6 Voting Application

The *voting* phase was used in *OOTW* at the end of the game show to allow the audience to vote for their favourite losing team member. The incentive to vote was the fact that the 'doomed planet' was set to explode and only one member from the losing team could be saved by the audience. The interface for this activity differed from the *pong* game insofar as the audience view (as seen by the cameras) was mixed into the broadcast stream during the playing of *pong*. Whereas during the *voting*, all the audience could see was the virtual environment moving as a function of their aggregate activity (see Figure 4.3 and Figure 4.6).

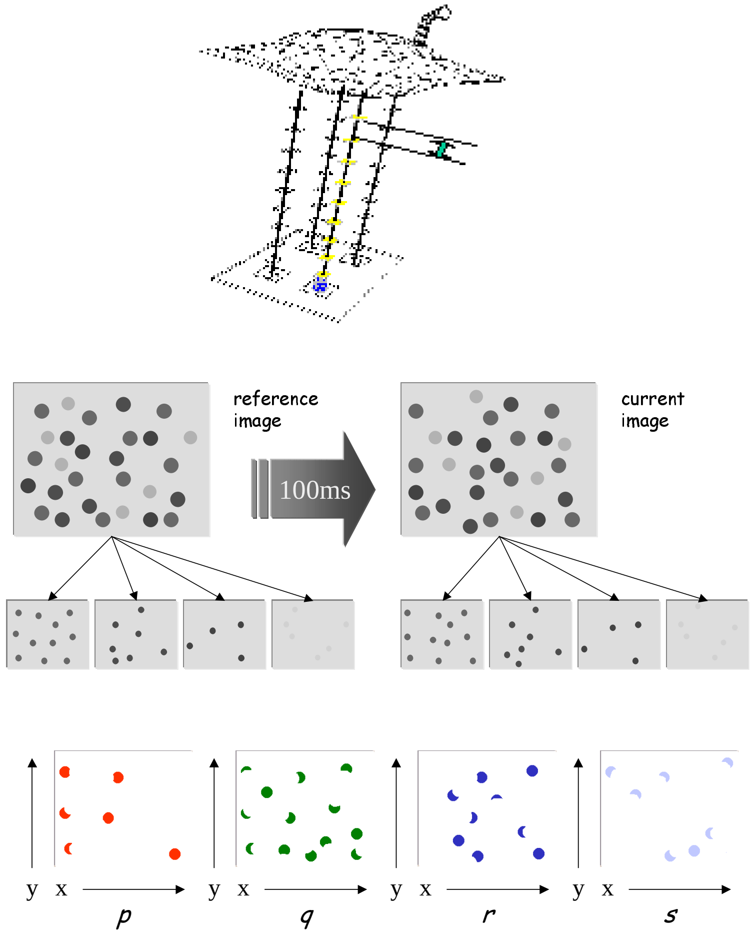

Figure 4.7: WobbleSpace II – Voting processing (top) shows the original storyboard design showing the four 'beams' used to elevate the performers; (middle) shows the capture and segmentation phase; (bottom) shows the calculation of unique movement through image subtraction.

The first issue to address when designing the *voting* application of WobbleSpace, was whether to make the activity a parallel or sequential one. In sequential voting, the audience is asked to vote in turn for each of the four losing team members. WobbleSpace is used to measure the overall activity in the camera view for each of the losers in turn. The one with the highest activity is picked as the winner. Basically this is a traditional 'clapometer', but based on visual activity rather that the volume of clapping noise.

Advantages:

Easy concept;

Each member of the audience is able to choose their personal favourite.

Disadvantage:

Possible problem with abstentions—will audience have enough time to identify a favourite during the game? This could be overcome by asking each loser to say why they should be saved (like balloon debate), though this would require more time.

Comments:

Feedback to the audience—how is this achieved? We originally viewed the space-ship as floating overhead. The losing team members would stand in a line and each would be raised up in turn by an amount proportional to the measured activity. Then the space ship could move down and pick up the highest (and nearest) loser.

In parallel voting, the audience would be able to root for one of the four losers, voting would take place all at the same time.

Advantages:

All the audience votes at the same time—it's quicker;

The audience is encouraged to vote as they are *competing* with other parts of the audience;

Allows the audience members to dynamically change allegiance.

Disadvantage:

May be more difficult, technically

Comments:

Feedback could be done as above, but the losers would be raised in parallel.

In the end, we chose to use the parallel voting mechanism. This entailed analysing the image (after segmentation into respective *p*, *q*, *r* and *s* values). Each value was summed to achieve a value *T*. This value was then used to find *p'*, *q'*, *r'* and *s'* by dividing *T* by *p*, *q*, *r* and *s* (to provide a normalised value for moving each of the performers. Finally, this normalised value was scaled (at run time) to calibrate the ascent rate into something that was perceivable as being fluid and that didn't either make the voting happen either too quickly or too slowly.

A final word is necessary on what the developers termed the 'production override feature' for the *voting*. We were sympathetic to potential effects of unpredictability in WobbleSpace and production. In particular, we were keen to provide a manual override in the unlikely event that the analysis process failed in order to not interrupt the performance flow. The intention was that the interface (as used by the developers in the Mezzanine) allowed us to manually generate values for the voting and override the analysis process based on what we saw from the camera views.

Fortunately, the manual override feature wasn't required. However, as with *pong*, there were a few seconds needed at run-time to set the scaling factor such that the *voting* application could operate with arbitrary numbers of audience members which again took ~5 seconds to finely adjust. Some aspects of this and other adjustments to the operation of WobbleSpace are further discussed in Chapter 6.

4.4.7 Summary

This section has described in detail the evolution of WobbleSpace. Outlining two versions of the software developed for different, but theoretically related, activities. We have focused upon WobbleSpace II for *OOTW* and have outlined the design influences and technical implementation undertaken in that development. The following section describes our experiences both subjectively through feedback sessions and meetings as well as statistically through the analysis of log files captured during the event.

4.5 Evaluation

Here, we outline the lessons learnt in the Green Room from the *OOTW* audience participation activities using both qualitative and quantitative measures. We have sought to analyse whether the experience was enjoyable, believable and worthy of inclusion as an activity[5].

Specifically though, we have sought to measure the amount of activity over both applications of WobbleSpace II (i.e. *pong* and *voting*) and seek to find whether there is a correlation between the perceived responses gained through feedback sessions and the descriptive statistics provided by analysing the log files. The objective analysis suggests that the audience certainly became engaged with the activity. However, subjectively at least, issues of 'believability' remain questionable.

4.5.1 Subjective Analysis (Audience Feedback Sessions & Post Event Meetings)

After each of the four performances of *OOTW*, the audience were invited to stay behind and contribute to an open and at times, frank, discussion about the whole event. Needless to say, these sessions proved invaluable in our learning about what had and what had worked not during the performance.

At these times, the WobbleSpace developers were integrated into the audience to both answer questions on WobbleSpace as well as take the opportunity to gather information from those who had just participated.

It would be fair to state that all four performances produced different results, although there were some consistencies throughout. For example, the interaction technique of waving cards was generally thought unsophisticated and more hi-tech mechanisms were often suggested without regard for the costs involved in producing 200 such devices! Evenso, the issue was raised by a number of people. Another, more positive, consistent view was that the *pong* experience was enjoyable. A fact supported by the log files analysis below in Section 4.5.2. In fact, one audience member from the first performance stated that "*pong* was the best part of the show".

Interestingly enough, from the log files, it was the first audience that seemed most active with both *pong* and the *voting*. However, apart from the final performance (discounted due to the low attendance), all the other audiences seemed to a greater or lesser degree to enjoy the experience. At times, the audience were heard to cheer when,

[5] The reader is encouraged to have viewed the accompanying *OOTW* video before proceeding beyond this point

for example, hitting the ball during *pong*. Another fact brought out by another member from the first audience was that *pong* 'bought the audiences attention'. This is an interesting observation. Especially when considering that during this 'warm-up' phase no 'entertainer' was present to cajole the audience. And, it was indeed the audience who cajoled and encouraged each other to participate.

Through observations of the video captured by the developers for off-line analysis, audience members can be seen standing and waving their cards in response to what they see. Of interest to our observations of this video is the fact that this behaviour was only evident during *pong*. We conjecture that this an be attributed to the fact that during this 'warm-up' activity, the broadcast projection combined the image the camera saw with the view of *pong* into a mixed reality view allowing the individual to get feedback on their individual contribution to the game (see Figure 4.8 below and the accompanying deliverable video).

Figure 4.8: OOTW *Audience Participation – two members of the audience can be seen on the left and right standing and waving their cards during show 1. Note that this is enlarged from a full frame with the camera join also visible in the middle. Also note the superimposed HLS pixels on the cards (see Section 4.4.4).*

The implication here is that by feeding back interaction to the audience, they are compelled to interact more by virtue of the fact that they recognise exactly what there current level of contribution is.

Of all the constructive feedback offered from the post *OOTW* discussions, the developers felt the issue of 'absence' most prevalent. It appeared that after 'engaging' with the content during the *pong* application, the audience weren't required again until the final voting phase of the performance. This fact troubled both the audience and performers (who were unable to see the audience throughout the performance). Although the performers were co-located with the audience in the theatre, they wore Head Mounted Displays and as such were shut away from the audience (visually at least). It transpired that this fact affected the performers activities as interaction and feedback is seen by many branches of theatre and television to be a key attribute to the performance as a whole. However, the audience felt neglected and powerless to 'cheer or jeer' at the performers and *OOTW* inhabitants throughout the remainder of the performance and felt that some abstract representation of themselves (through WobbleSpace) could have ameliorated this discontent.

Interestingly enough, this endorses the one of the secondary objectives under research with the WobbleSpace programme—i.e. 'to what extent do the crowd feel part of the virtual environment'. Deeper discussion on this factor is deferred to Deliverable D6.2. However, it was clear that the audience did indeed feel that they were not as involved with the majority of the game show as they had expected (following *pong*) and suggested that we could have used the technology to have given them a greater degree of presence throughout the remainder of the show, rather than just inviting them to 'vote' at the end.

4.5.2 Objective Analysis (Log Files and Descriptive Statistics)

Taking a step back from the discussions and feedback, we now turn our attentions to the statistical analyses from the log files captured as both the *pong* and *voting* applications in *OOTW* took place.

Of note is the fact that for this section, we are only drawing upon two of the four performances. This is due to the reason that session three's logs were corrupt and session four was discounted because of the low attendance (due to the performance being late Sunday and colliding with another event at the conference). However, the two performances that are analysed account for ~60–70% of the total number of audience members.

Here, we are primarily concerned with measuring the 'amount of activity' in each show over both the *pong* and *voting* applications. We will highlight disparities between the activities and seek to reason about the observable activity patterns.

Discrepancies in data capture

As mentioned earlier in Section 4.3.2, noise interferes with machine vision techniques in all application domains ranging from industrial manufacturing to entertainment. WobbleSpace suffered from noise. Although minimal, the amount of noise in the scene did indeed interfere with the data captured. For example, emergency lighting in the theatre, which obviously couldn't be turned off, clashed against one of the HLS values that the image engine was looking for. Additionally, during show 1, all seats in the theatre had the coloured card placed on for the audience members to pick up when they arrived. Although WobbleSpace is designed to eradicate sustained (stationary) colours, disparities between samples did introduce some noise into the results. However, these were minimal, and the developers are satisfied with the analysis presented below.

Analysis of overall activity rate

Activity is a 'measurement of exertion' (Concise Oxford Dictionary, 1996). We sought to measure activity in two ways; as an overall rate (i.e. summing p, q, r and s) and individually. This section is concerned with the observable behaviour gathered as a function of the summed (overall) activity for both *pong* and *voting*.

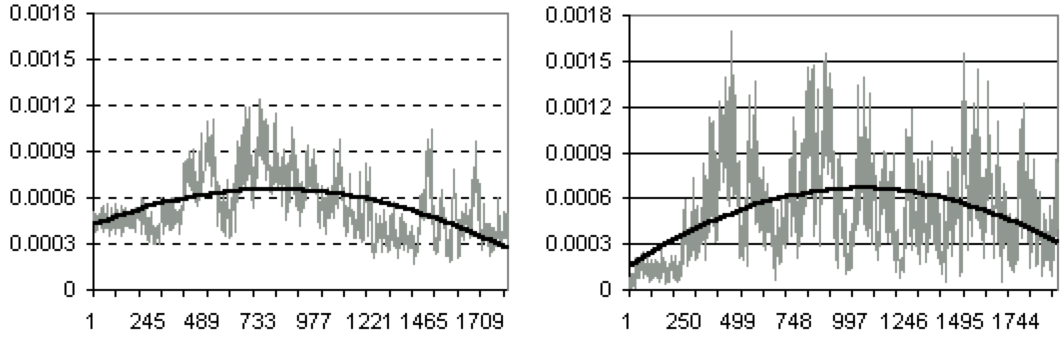

Figure 4.9: WobbleSpace II – Overall activity rates for pong between show 1 (left) and show 2 (right).

Figure 4.9 shows the activity observed from two shows of *pong*. The x-axis represents the number of samples taken (where the average sample rate was at 10Hz) with the y-axis showing the proportionate amount of *unique* activity as a function of the whole image. For example, the mean value for show 2 across the sample equates to 0.00052 (SD = 0.00031) which translates to ~229 unique pixels per sample.

Show 1's mean was 0.00054 (SD = 0.00019) which equates to ~242 unique pixels per sample. What does this tell us? Although, not much can be gleaned from the mean values as the sample sizes (and audience sizes) were different, we can infer something from the standard deviation. This shows that activity (for show 2) was spread over a wider range of values indicating that activity was more sporadic. However, inference on overall activity can further be gleaned by applying a second order polynomial trendline to the sample sets as shown in Figure 4.9 (and Figure 4.10 below).

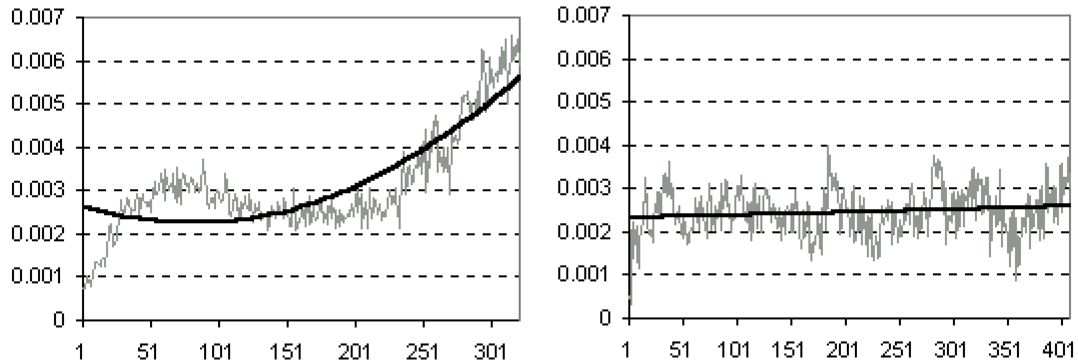

Figure 4.10: WobbleSpace II – Overall activity rates for voting between show 1 (left) and show 2 (right).

From the trend line in Figure 4.9 we can see that in both cases of *pong* the logged activity matches up to that observed through watching the video footage. It was observed and supported by the logs, that the audience took some time to get into the *pong* activity. This seemed to stem from some confusion over exactly *how* to interact with the screen. However, after a period of settling down (in the order of 45–60 seconds), the audience managed to interact more fluidly with the *pong* screen, and indeed appeared to enjoy the experience. Towards the end however, they became bored and activity is seen to taper off.

Figure 4.10 shows a similar analysis applied to the *voting* application. Here we see a slightly different pattern of activity. For example, as expected, the overall mean activity for both shows during the *voting* was higher than during the *pong* game.

	Application	Mean	SD	Max	Min	Unique pixels/sample
SHOW 1	Pong	0.00055	0.00020	0.00124	0.00017	242.8
	Voting	0.00313	0.00117	0.00658	0.00074	1382.4
SHOW 2	*Pong*	0.00052	0.00031	0.00169	0.00001	229.6
	Voting	0.00245	0.00054	0.00396	0.00034	1084.3

Table 4.1: WobbleSpace II – Statistical values from shows 1 and 2

Although both graphs from Figure 4.10 exhibit different profiles, they both show a sustained, and in the first case, increased, level of engagement. In particular, show 1, whose audience were actually more enthusiastic (through cheering etc.), shows a marked increase in activity over the ~35 seconds duration of the voting. This can be accounted for by analysing the individual activity rates as shown in Figure 4.11 below.

Analysis of individual activity rate

The graphs in Figure 4.11 show the normalised activity rates for the individual colours (i.e. *p*, *q*, *r* and *s*) for the *voting* application of WobbleSpace over both shows 1 and 2. They illustrate the individual activity rates, per colour, over the duration of voting and illustrate some interesting emergent activities.

In particular, they appear to illustrate that the rate of voting for the losing team seemed to increase over time. Or, put another way, once the viewers supporting the winning 'inhabitant' were comfortable that they had won, the overall 'energy' expended voting tailed off. We could also consider the possibility that certain members of the audience changed sides half way through the voting. And, although evident, that was indeed an exception and not the rule. In the main, as evidenced from analysing the video footage, the viewers generally supported one member throughout the voting phase.

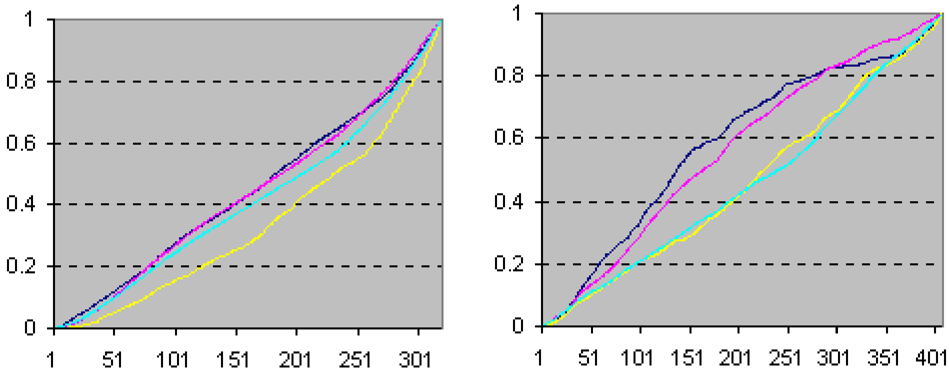

Figure 4.11: WobbleSpace II – Individual activity rates for voting between show 1 (left) and show 2 (right).

Therefore, it is our conclusion from Figure 4.11, that parallel voting seemed to engender a degree of viewer competition which certainly wouldn't have been present under a sequential mechanism.

Analysis of frequency distributions

Our final analysis is concerned with looking at the frequency distribution of activity throughout both shows 1 and 2 and comparing the distribution of activity between both the *pong* and *voting* applications.

Figure 4.12 below illustrates the findings. Both samples were equally distributed into 20 buckets. The intention is to show the range of normalised activity whereby the peak (up the y-axis) shows where most activity occurred and the range (x-axis) indicates the amount of increasing activity from left to right.

Thus, turning our attentions initially to *OOTW* show 1 (Figure 4.12, left), we can see that both applications (*pong* and *voting*) are negatively skewed indicating that most of the activity was concentrated towards the lower end of the activity scale. However, both shows (during *pong*) exhibit a similar trend, a relatively sharp increase in the overall activity, tailing off slowly. This would indicate that once the viewers had got started, they operated at an optimum rate until they slowly became bored of what they were doing. Whereafter, activity is seen to diminish.

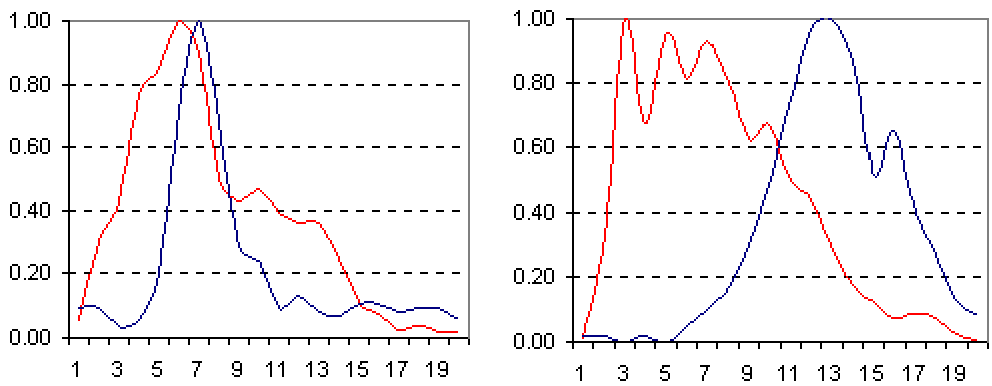

Figure 4.12: WobbleSpace II – Frequency distribution between show 1 (left) and show 2 (right). Note that pong is illustrated through the red line and the voting activity through the blue line. Bucket sizes for show 1 were 0.00005 (for pong) and 0.00029 (for voting) with show 2 being 0.00008 (for pong) and 0.00018 (for voting).

Show 2, on the other hand, shows a different behaviour (between *pong* and *voting*). *Pong* exhibits a similar profile to that shown in show 1 – sharp increase, then taper, but the *voting* exhibits a more even (i.e. normal) distribution about the mean. It is difficult to account for this discrepancy objectively and no direct conclusions can be inferred at the present time.

4.5.3 Summary

This section has presented the qualitative and quantitative evaluation of WobbleSpace II at *OOTW*. Primarily, we have sought to correlate observable behaviour in the video analysis against the subjective feedback sessions and log file analyses.

From these data sources, a number of conclusions have been presented:

- On the whole, the audience enjoyed the experience

- The audience were at times, frustrated with the 'primitive' interaction capabilities
- The audience seemed to engage in competitive behaviour when voting
- Patterns of activity in the logs matched up with observable behaviour on video
- The audience wanted more involvement with the game show

4.6 Conclusions and Future Work

This chapter has covered the development, design, execution and evaluation of WobbleSpace I and II. During *OOTW*, we sought to simulate a limited form of Inhabited TV *viewer* participation to gauge the levels of response such a future audience may both exhibit and expect. And, although a theatre based experiment, in the case of *OOTW*, it has uncovered a number of interesting requirements for future digital Inhabited TV productions. First, the audience want to be involved. They firmly expressed this on a number of occasions. It is insufficient to merely invite them in at strategic points in the narrative in order to satisfy their need to participate. A degree of 'presence' (or rather, absence?) seemed to be a prevalent issue. This was highlighted by not only the *viewers* in the theatre, but also by the performers in the virtual environment.

Additionally, designing for a participatory experience needs to take account of real-world societal situations where audience participation is considered an intrinsic component in the success of a performance. Bilson's heuristics for audience participation were adopted for the *OOTW* event, but it would be fair to say that although beneficial, a more persistent form of audience presence, expression and engagement would have certainly helped the production from both the viewers and performers perspectives.

However, analysing the video did show a willingness to 'interact' with the production, a process of 'buying-in' the audience was definitely seen as a positive gesture. Indeed, although subject to noise, which was apparent to the audience, they (the audience) persisted with the interaction just in case their activity was picked up by the camera and image processing.

Finally, the audience did at times cheer and clap, indicating pleasure at our attempts to involve them. And, although WobbleSpace II was hindered by the less than ideal theatre conditions, this proved a willingness to participate. Therefore, it is our suggestion that future designers of i-TV game shows would be well advised to consider audience participation *early* in design and in particular, how to engage the audience with the performers *throughout* the production.

We conclude with the findings we consider most important:

1. Engage the audience throughout the production
2. Connect the audience and performers together
3. Encourage competitive activity (e.g. parallel voting)
4. Provide feedback mechanisms recognising the (individual) interaction
5. Adopt heuristics from real world participatory activities

Chapter Five: Network and User Analysis for *Out Of This World*

Chris Greenhalgh
Department of Computer Science, The University of Nottingham, Nottingham, UK

5.1 Introduction

This analysis is based on data from the four performances of *Out Of This World*, staged on the 5th and 6th September 1998. In particular, we make use of the following:

- logs of network traffic data captured using the tcpdump tool; and
- log files from the world manager process, which records the initiation of phases in the game and also records the presence and movement in the virtual world of all users.

To determine the exact timings of the performances we made use of video recordings captured from a monitoring machine, plus clock synchronization information from the PLOD distribution platform.

We begin by examining the network traffic from a particular show (show 2) in detail. In particular, we:

- explore the network traffic from that show, identifying the key types of traffic and the form of the performance's bandwidth requirements; and
- consider the impact of show phases on the network traffic; and
- compare the bandwidth requirements of different traffic sources according to their level of participation (host, captain, inhabitant, camera);

We then move from direct consideration of network traffic to consider user activity (in particular movement and speaking) during the shows. In particular, we:

- analyze the overall rates of moving and speaking as seen by the network, and relate these to previous studies;
- characterize the impact on speaking and movement of the different levels of participation employed in *OOTW*;
- compare the levels of activity of the two teams (one with male members and other with female members);
- compare the levels of activity in different shows; and
- explore the apparent effects of phases and activity constraints on user activity.

5.2 Show 2 Network Traffic

In this section we explore the network traffic from show 2. Figure 5.1 shows the total bandwidth for this show, calculated at one minute intervals. It also shows the top-level traffic breakdown into multicast UDP traffic, unicast UDP traffic and TCP traffic. The traffic trace begins at 20:03:08 on the 5th September 1998, just before the world and all users are restarted for the show. The actual broadcast from the world runs from 20:43:02 (time 2394) to 21:26:52 (time 5024)—traffic is shown for a further 5 minutes after the end of the show.

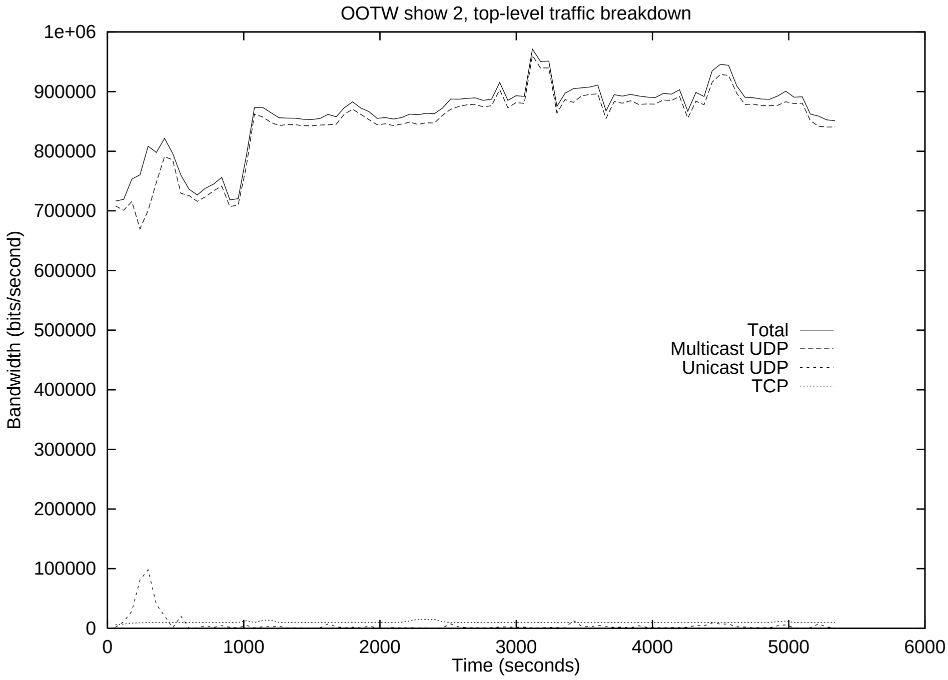

Figure 5.1: Top-level traffic breakdown for show 2.

The table below summarizes the traffic breakdown.

Type	*bytes*	*%*	*rate bit/s*
UDP-MC	564093044	98.2	845083.2
TCP	6759358	1.2	10126.4
UDP-UC	3737055	0.7	5598.6
Other	9380	0.0	14.7
Total	574598837	100.0	860822.2

We observe from the figure and the table that:

- The network traffic is consistent throughout, at around 860 kbits/second.

- The vast majority of this traffic (98.2%) is multicast UDP traffic. This in turn comprises audio, video and application components, which are explored below.
- The unicast UDP traffic is concentrated at the start of the session, between time 0 and time 600; this corresponds to starting up the world and the users, when unicast is used for state transfers. There are smaller bursts of unicast UDP traffic throughout the session; unicast is also used for ensuring reliability of multicasting and for various direct interactions (e.g. applying management constraints to users).
- There is an almost constant background level of TCP traffic at around 10 kbits/second. This is considered below.

The following subsections consider the multicast, unicast and TCP traffic components in more detail.

5.2.1 Multicast Traffic Components

Using destination port numbers we can identify the main traffic types within the total multicast traffic. Their contributions are summarized in the table below and plotting in Figure 5.2.

Type	*bytes*	*%*	*rate bit/s*
Audio	523160128	91.0	783760.5
Video	22709952	4.0	34804.5
CVE	18219904	3.2	27295.7
Other MC	3060	0.0	4.7

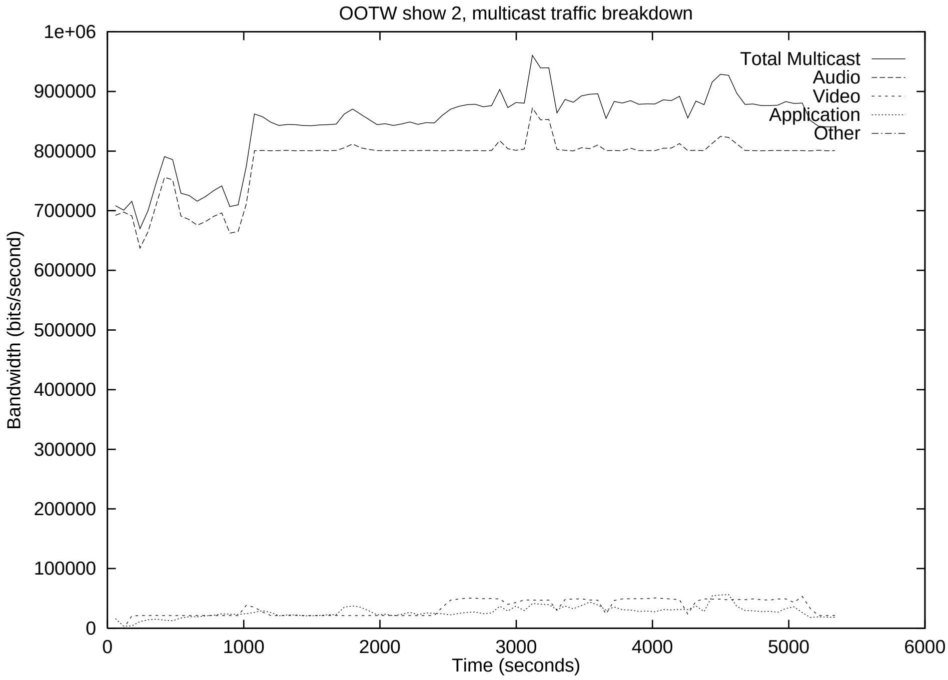

Figure 5.2: Multicast traffic components for show 2

Considering each component in turn:

- After time 1000 (before the start of the show) the audio traffic level is very stable at about 800 kbits/second throughout the show. There are small bursts of additional traffic, for example around time 3200 and 4500. The background audio level corresponds to real-time audio from the 11 users (host, two captains, eight inhabitants). This level is so consistent because the normal silence detection mechanism was disabled. This was done on site to avoid audible artefacts from the operation of the mechanism. Consequently, every user sends audio traffic continuously. The localized bursts of traffic are sound effects generated during the games.
- The video traffic moves between two steady levels of about 21 kbits/second and 49 kbits/second. There is a single video stream for the host. This uses JPEG compression for each frame, and consequently the resulting traffic is greater for images which contain more complexity. In this case, the lower bandwidth corresponds to times when the host is not in front of the camera (e.g. before the event starts around time 2400), while the higher bandwidth is when the host is present, i.e. for most of the game. During the end-of-game score review the host screen is not visible to the inhabitants and the host appears to move away from the camera, accounting for the dips around time 2900, 3300, 3700 and 4200.
- The application (CVE) traffic accounts for changes in the virtual world, such as: users moving, the appearance and disappearance of speech balloons, and the activity of objects within the worlds (e.g. the space frogs). This has a fairly

consistent background level of about 20 kbits/second, rising to a maximum of 56 kbits/second (one minute average) around time 4500 (the race). We consider the relationship of this variation to the phase structure in more detail below.

Multicast traffic and show phase

Figure 5.3 shows the four components of the multicast traffic which are most affected by the progression of the show itself for the main period of the show. For reference, the main stages of the show are as follows:

- Pre-show football match, time 1700–1900;
- Start of show, time 2400;
- Football match during introduction, time 2800–2890;
- Move to arena 1, time 2945–3020;
- Space frog game, time 3050–3240;
- Move to arena 2, time 3290–3370;
- Falling fish game, time 3400–3600;
- Move to arena 3, time 3670–3760;
- Quiz game, time 3770–4200;
- Move to arena 4, time 4260–4350;
- Race, time 4370–4570;
- WobbleSpace interaction, time 4895–4955;
- End of show, time 5025.

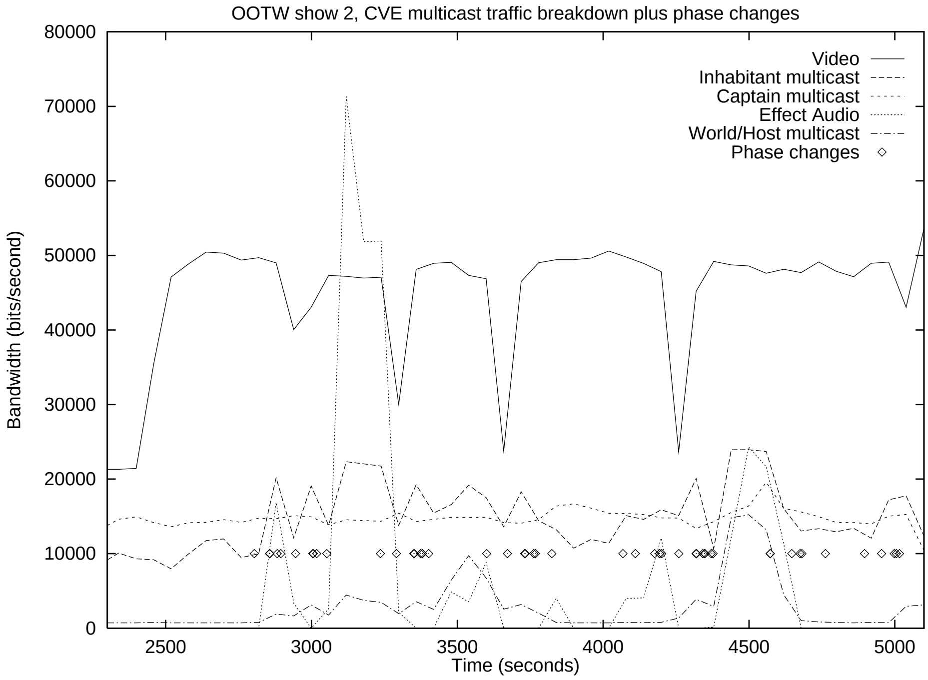

Figure 5.3: Breakdown of CVE multicast traffic plus sound effect audio, with phase changes shown for comparison.

For the traffic components shown we note the following.

- The video of host, as noted above, increases when the host is in front of the camera—at the start of the show, and decreases again during each end-of-game debriefing (times 2945, 3290, 3670 and 4260).
- The sound effect audio occurs during the football game (the ball "booings" when it is hit), the space frog game (they "croak", and "squeal" when impaled), the fish game (fish "chime" when harvested), the quiz (the rings give a fanfare when passed) and the race (the cones "crash" when hit). The correspondences to phases are thus very clear.
- The team captains produce a very consistent level of traffic at around 15 kbits/second, total. This is dominated by the continuous stream of position updates which are generated from the use of magnetic trackers to control their embodiments (see the user movement analysis in Section 5.3 for more details).

The remaining CVE application traffic are also shown in Figure 5.4 at a finer level of temporal detail (10 second intervals).

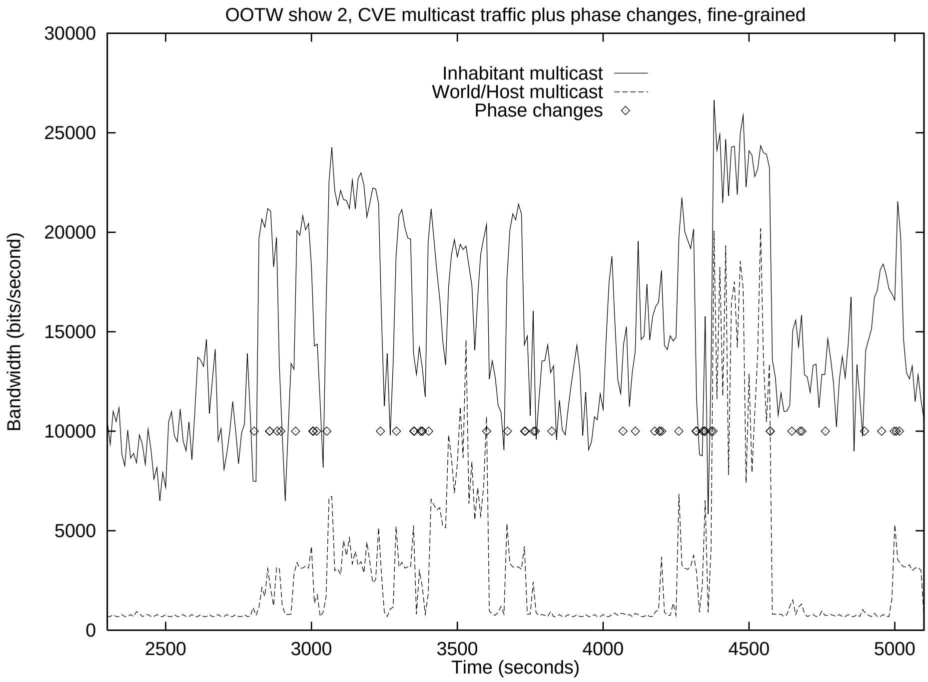

Figure 5.4: CVE multicast traffic with phase changes shown for comparison.

We note that:

- The world/host multicast traffic has a constant background level around 700 bits/second; this is keep-alive messages for the world to show that it has not crashed. There are peaks in activity during the games corresponding to the activity of world objects such as the football, the space frogs, the platforms (in the fish game) and the cars in the final race. Again the relationship to phases is direct and simple. There are also blocks of traffic during movements between arenas corresponding to the movement of the host's embodiment.
- The inhabitant application multicast traffic exhibits quite a fine-grained relationship to the phases of the game, although the overall variation is still relatively limited, e.g. less than 10 kbits/second at times in the relatively controlled entries to and exits from arenas (e.g. times 2900 and 3050), rising to about 25 kbits/second during the race (time 4370–4570). However most phases exhibit substantial variability, for example during the quiz (time 3770–4200) the inhabitant application bandwidth ranges from 9 to 19.5 kbits/second. In the later section which examines user activity we will look again at the relationship between phases and user activity (and its network implications).

5.2.2 Unicast Traffic Components

We have already noted that the unicast UDP traffic is concentrated at the start of the period, before time 600, where we would expect state transfers to be taking place. The unicast UDP traffic in MASSIVE-2 has three roles: state transfers, ensuring multicast reliability, and point-to-point communication (e.g. the world manager imposing constraints on users and objects).

In order to estimate the amount of reliability-related traffic we consider separately:

- unicast traffic to and from the various node manager processes, which handle world and object name resolution and to monitor clock offsets between machines on the network (for post-event compensation and analysis);
- the unicast traffic which flows to and from the world host, which should include world state transfers plus all management traffic (and associated reliability-related traffic); and
- unicast traffic between other machines, which will include initial state transfers (for user embodiments), but only reliability-related traffic after this.

The breakdown is shown in Figure 5.5 and in the table below. The application multicast traffic is also shown for comparison, as are the moments at which phase changes are triggered (shown as points on the graph—vertical position has no inherent meaning for these).

Type	*bytes*	*%*	*rate bit/s*
MC CVE	18219904	3.2	27295.7
World	2340236	0.4	3506.0
Rest	925035	0.2	1385.8
NodeMgr	471784	0.1	706.8
Total	3737055	0.7	5598.6

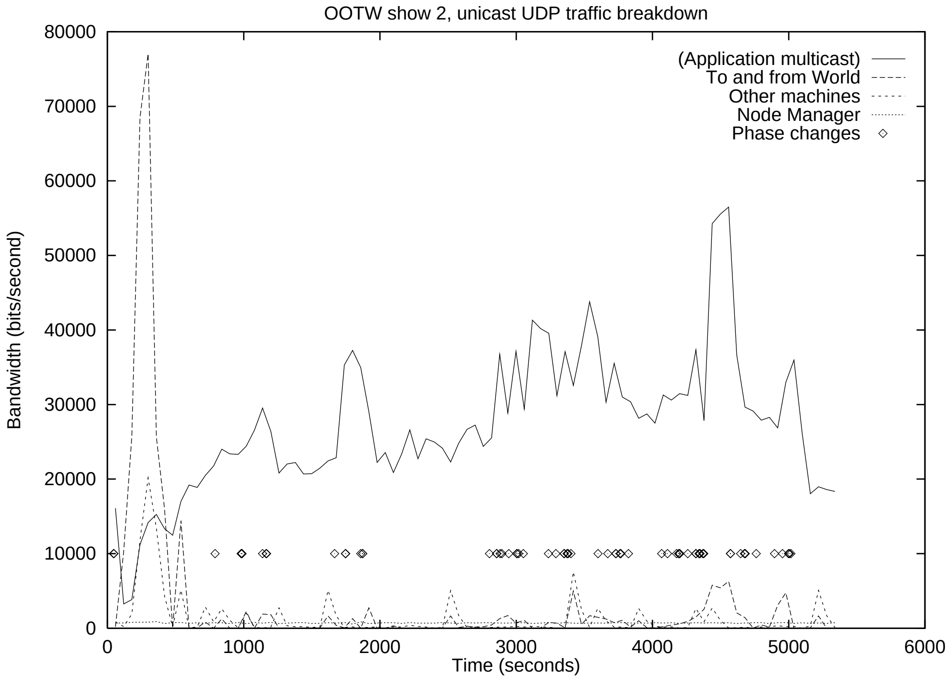

Figure 5.5: Traffic breakdown for unicast UDP traffic, show 2 (including application multicast traffic and show phase changes for comparison).

We make the following observations:

- State transfers appear to be (and we would expect to be) concentrated in the first 600 seconds. This period accounts for 60% of the total UDP unicast traffic.
- There are some clear examples of phase-related unicast traffic, between the world machine and other machines, for example around times 1200, 1900, 2900.
- There is also clearly some residual reliability-related traffic. This appears to be of the order of 10% of the unicast traffic, or equivalently less than 1% of the corresponding multicast traffic. It is quite peaky in character, though it bears no clear relation to variations in the multicast traffic.
- Node manager traffic, primarily timing checks, accounts for around 12% of the total unicast traffic. It runs consistently throughout at around 700 bits/second.

5.2.3 TCP Traffic Components

We have already seen from the top-level breakdown that the TCP traffic is a very consistent background level of around 10 kbits/second. If we examine this in more detail we can identify the following components:

- Remote login traffic, used by the system administrator to monitor the progress of the event. The constant rate indicates that they are receiving a relatively constant rate flow of status information. There is one additional burst of remote activity around time 1000.

- Process watchdog, which was a TCP-based program to monitor the liveness of the various programs comprising the complete distributed system. This was intended to assist in detecting and diagnosing process failures, though in fact the software was stable throughout. This is essentially a continuous flow of status messages from all active processes.
- A block of TCP traffic around time 2300 which is the network communication for interfacing WobbleSpace to the pong game in the opening sequence.
- WobbleSpace events, around time 4900, which is the network communication interfacing WobbleSpace to the virtual world to drive the inhabitants up the tubes to spaceship.
- A little video membership monitor traffic around time 100–400 as new user processes start up and register with the video service.

This traffic breakdown is shown in Figure 5.6 and summarised in the table below. This accounts for all of the observed network traffic.

Type	*Bytes*	*%*	*rate bit/s*
Login	4400621	0.8	6592.7
Watchdog	2148560	0.4	3218.8
Other TCP	154105	0.0	266.8
WobbleSpace	29836	0.0	47.9
MembMon	7548	0.0	11.8

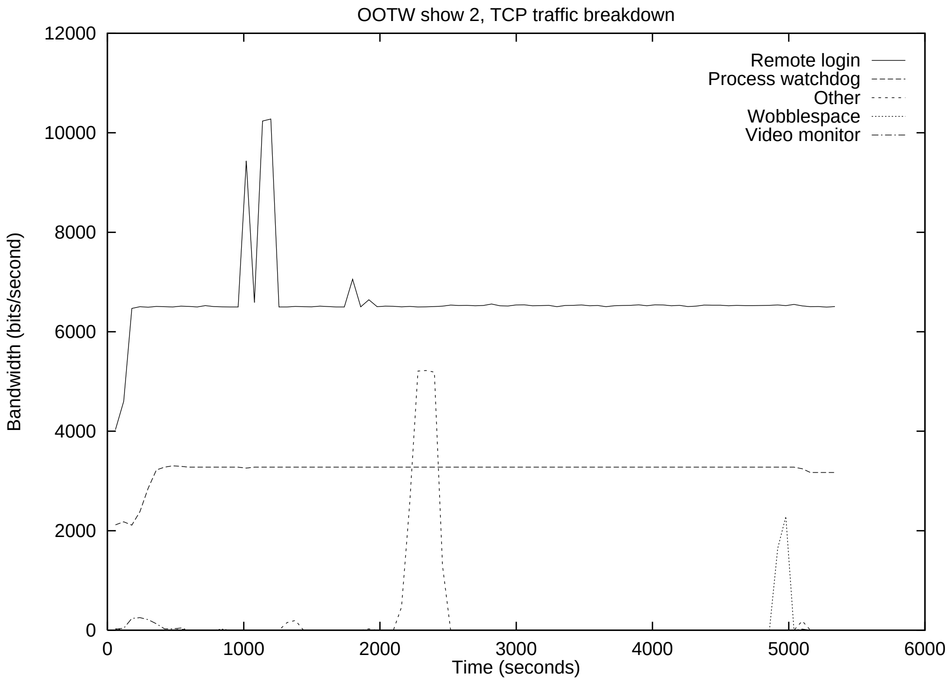

Figure 5.6: Breakdown of TCP traffic, show 2.

5.2.4 Network Traffic and Levels of Participation

In the previous section we explored the network traffic from show 2, and showed how it breaks down into multicast UDP, unicast UDP and TCP components. We further showed how these break down into their constituent activities. In the case of the application component of the multicast traffic we have shown how this varies (to a limited extent) with the various phases of the show as it progresses.

In this section we relate network traffic to the different levels of participation used in *OOTW*; these were:

- Host, the show compere, embodied using live audio plus video (the only participant with video);
- Team captain, using an immersive tracking-based interface with live audio;
- Inhabitant, using a desktop joystick-controlled interface with live audio; and
- Camera, using a desktop, mouse-based, interface, not normally visible in the virtual world and having no live audio.

Firstly, there is simple relationship between level of participation and media, which in turn has a direct impact on the network. Specifically:

- Only the host has a video stream, which consumes around 49 kbits/second when they are on-screen (around 20 kbits/second when they are not). So the host's bandwidth requirements are of course 49 kbits/second greater than an equivalent non-video participant.

- The host, captains and inhabitants all have real-time audio which contributes 72.8 Kbits/second of audio network data. Recall, that in *OOTW* silence detection was disabled to improve the perceived quality and reliability of the audio channel, so all potential speakers send audio data all of the time.

The situation for virtual world (application) data is potentially more complicated. It is not readily possible to distinguish the traffic from the Host (it appears within the multicast traffic from the world as a whole). It is, however, possible to identify the application traffic for each captain and inhabitant. For the broadcast part of the show (from time 2394 to time 5024) we find the following (average) levels of traffic:

Type	*bytes*	*%*	*rate bit/s*
CaptainA	2560924	0.8	7587.9
CaptainB	2474900	0.8	7333.0
Host/world	1002292	0.3	2969.8
Inhab4A	802916	0.3	2379.0
Inhab1B	677232	0.2	2006.6
Inhab3B	631068	0.2	1869.8
Inhab3A	599076	0.2	1775.0
Inhab2B	532292	0.2	1577.2
Inhab2A	481068	0.2	1425.4
Inhab1A	420252	0.1	1245.2
Inhab4B	415292	0.1	1230.5
Camera2	157756	0.1	467.4
Camera4	157680	0.1	467.2
Camera3	98268	0.0	291.2
Camera1	84616	0.0	250.7
Camera5	80308	0.0	237.9

We observe that:

- The team captains generate around 7500 bits/second on average, over six times the average for the inhabitants. This is as we might expect, since each captain's embodiment comprises four components (head, body, left hand, right hand), and being tracked they are likely to move more (as far as the system is concerned).

- The combined host and world traffic is still only a little more than a single inhabitant's traffic.
- The inhabitants have an average bandwidth (in this show) of around 1800 bits/second. There is considerable variation between inhabitants, from 1230 bits/second to 2379 bits/second, a range of 1149 bits/second. This comprises movement updates and the appearance and disappearance of their speech balloon. Inter-user variations in activity are explored in more detail later in this analysis.
- The cameras (four used for the broadcast plus one additional observation machine for diagnostics and logging) have an average bandwidth of only 340 bits/second (maximum for one camera of 470 bits/second). Even the larger of these is almost four times less than an average inhabitant; the average is over five times less. This figure is lower because the network update rate of cameras was reduced significantly compared to inhabitants. Consequently, the bandwidth requirement approaches that required to just signal that the process is still alive. The system might be further developed to support a purely observing form of participation; in this case the cameras would generate no additional multicast network bandwidth.

From the above we can predict an average bandwidth requirement for each level of participation, including all media:

Level	*Bits/s*	*Components*
Host	125000	audio, video, application
Captain	80300	audio, movement (4-part, tracked)
Inhabitant	74600	audio, movement (1-part, joystick)
Camera	340	keep-alive, degraded movement

Clearly, for a finite bandwidth, the level of participation of a user can have a profound effect. In particular, it should (in principle) be possible to add many more camera-like disembodied observers than active participants.

There are three qualifications to make here:

- These figures are without silence detection (which was the case for all four shows), however another situation which allowed its use could significantly reduce the impact of using audio (in the average case, but see also the next point).

- The application/movement figures depend on the actual amount of movement occurring, and particular users, application or interfaces could give rise to much higher levels of updates - see the further consideration of this kind of issue in the later section on user activity modeling.
- Adding any observer (e.g. a camera) may not add much or anything directly to the multicast traffic, but it does add extra load to the multicast reliability mechanisms: additional traffic (in this case unicast UDP) would be generated by those users when multicast packets were lost. This must be taken into consideration when determining how many observing processes might be supported.

5.3 User Activity: Movement Rates

We now move from direct consideration of network traffic to consider user activity and in particular the amount of movement and speaking undertaken by users during the four shows. In this and the following sections we consider the overall rates of activity across all participants and shows and also seek to break this analysis down to explore possible systematic variations due to levels of participation, team, show and phase.

We begin by considering participant movement. Each participant (host, inhabitant and team captain) has an embodiment within the virtual world. We begin this user activity analysis by considering the movement of each user's embodiment. From a network perspective this is interesting because movement update messages need only be sent when an embodiment is moving; no traffic is needed while it is stationary. In *OOTW* an embodiment can move because:

- the user whom it represents is causing to move, e.g. an inhabitant moving their joystick or a team captain moving a tracker;
- the management system may be imposing (changing) movement constraints on the embodiment which cause it to move; or
- there may be noise in the system (e.g. the team captain's magnetic trackers) which cause the system to see apparent movement.

5.3.1 Overall Movement Rates

For all participants in all shows we find that the mean level of apparent movement is 59.7% (SD 25.4%), i.e. on average each participant spends 59.7% apparently moving as viewed by the system (and generating update events). This is much higher rate than we have found in previous trials such as ITW (19.6%), COVEN dVS (15.6) or COVEN DIVE (26.3%). Figure 5.7 shows the cumulative distribution of movement rates for all participants in all shows.

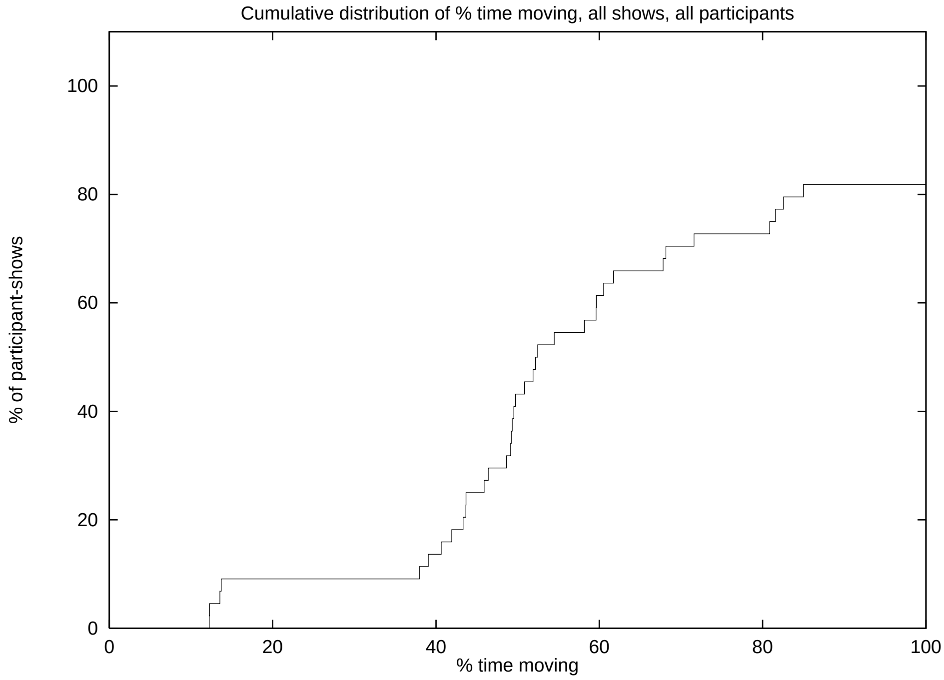

Figure 5.7: Cumulative distribution of movement rates.

It is apparent from the figure that the distribution is far from normal, with four participants giving very low movement rates (around 13%), and eight giving maximal movement rates (100%). There is a more reasonable spread of values in the range 38–85%.

5.3.2 Movement and Levels of Participation

To understand this distribution better we show the distributions for host, team captains and inhabitants separately in Figure 5.8.

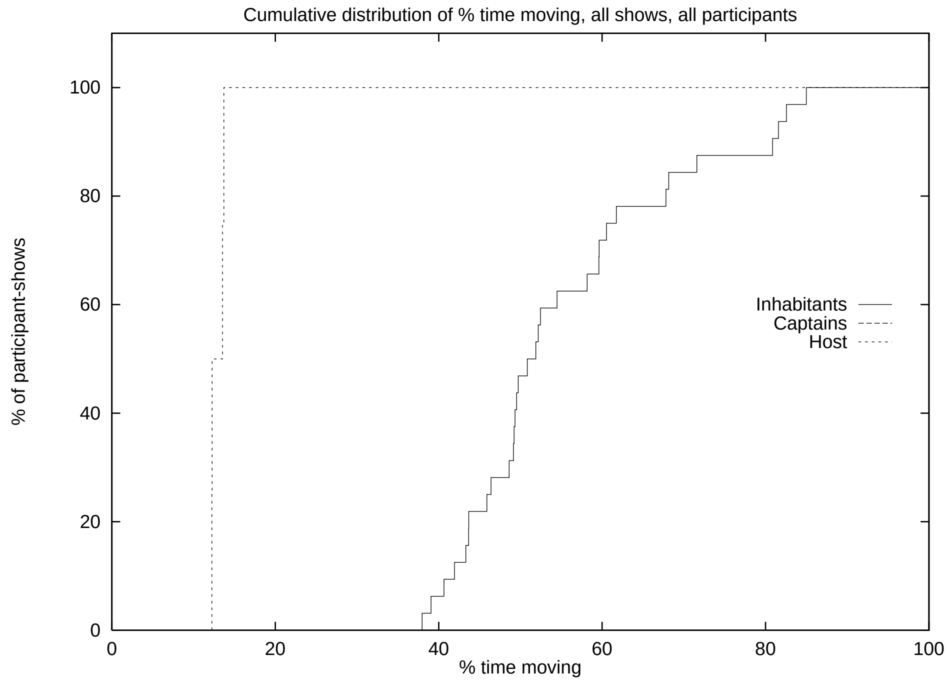

Figure 5.8: Cumulative distributions of movement rates by level of participation.

We find that the distinction between levels of participation account directly for the tri-modal form of this distribution:

- The host has a consistently low movement rate, average 12.9%. This is because the host is not directly controlled in the way that the other embodiments are, rather it is moved and positioned solely under the control of the event management system, to position it for different phases of the shows.
- The team captains have consistently maximal movement rates, all 100%. There is sufficient noise in the magnetic tracking system (as used) that the team captains always appear to be moving, as viewed by the system, even if they are ‘standing still’. Consequently the team captain generate position update events continually, throughout the event.
- The inhabitants account for the central region of the distribution, with a mean of 55.5%, standard deviation of 13.2% and a range of 38% to 85%. This is still much higher than observed in previous trials, and the reasons for this are considered next.

5.3.3 Movement and the Management Interface

Inhabitants’ embodiments can be moving for one of two reasons:

- the user is moving the joystick; or
- the event management system is moving them though the movement constraints mechanism.

In order to differentiate these two effects we have extended the user interface and movement messages to distinguish between movement events caused by the user, and movement events caused by constraints. This distinction is only useful for the inhabitants: the host is exclusively moved by constraints while the team captains always appear to be moving themselves (due to variations in tracker positions). For the inhabitants, if we ignore movement caused only by constraints, then we find that the mean rate of movement is 44.5%, with a standard deviation of 15.1% and a range of 24.2% to 82.2%. The cumulative distribution is shown in Figure 5.9, with the original distribution shown for comparison.

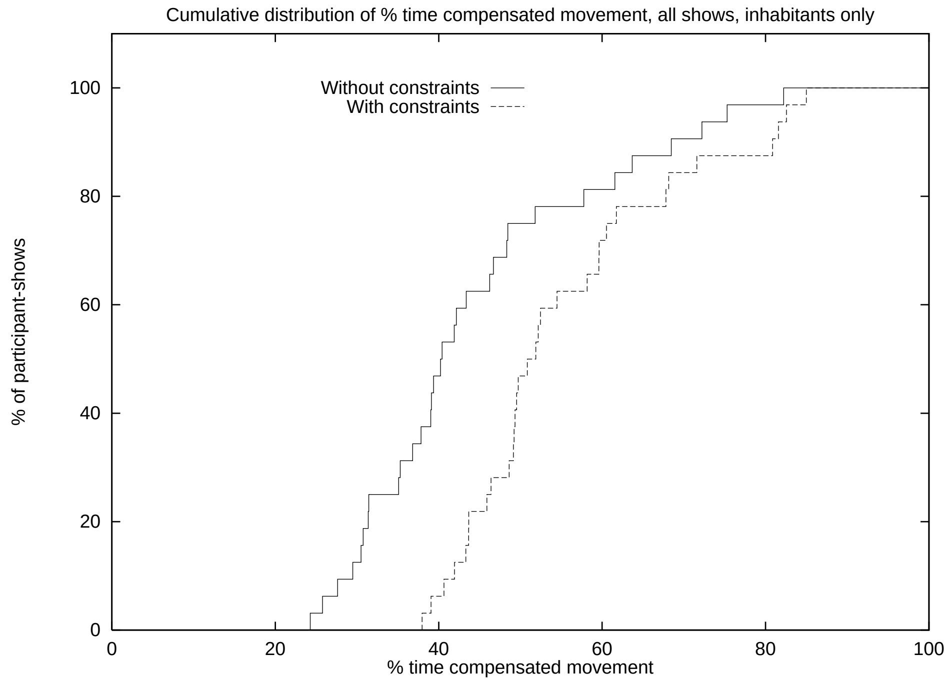

Figure 5.9: Cumulative distributions of movement rates for inhabitants, ignoring movement due to constraints.

So, the imposition of constraints has produced on average 10% additional apparent movement, and corresponding network traffic. This influence ranges from 13.8% additional movement in the case of the least active inhabitant and only 2.8% in the case of the most active inhabitant.

This compensated level of movement should now reflect the actual actions of the users themselves. In comparing this to the much lower levels from previous trials we identify the following differences:

- The most important distinction is the kind of task being undertaken within the virtual environment. The previous three sets of trials were all essential tele-conferencing applications, with some additional themed content. In *OOTW*, however, movement is a critical part of all of the games.
- The inhabitants in *OOTW* had access to no other applications or input devices, and so had no alternative activity within the computer system other than engaging in the show. They did this through moving and speaking. The previous trials had more complex interfaces and the potential for additional

parallel activities (e.g. in other applications on the user's machine), in addition to activity within the virtual world.

- The interface for inhabitants in *OOTW* is a single joystick (and headset for audio). The previous trials used primarily mouse-based interfaces, with some use of the keyboard. The use of the joystick may make movement easier for longer periods.

Further research would be required to explore the relative importance of these possible influences.

5.3.4 Movement and Teams

The games had two teams, team A normally had female members, while team B normally had male members (there were one or two instances of team swapping). If we compare the movement rates for the two teams (excluding movement solely due to constraints) we find:

- team A (female) had a mean movement rate over all shows of 40.2%, standard deviation 14.0% and range of 24.3% to 82.2%;
- team B (male) had a mean movement rate of 48.8%, standard deviation of 15.3% and range of 27.6% to 75.3%.

The cumulative distributions are shown in Figure 5.10.

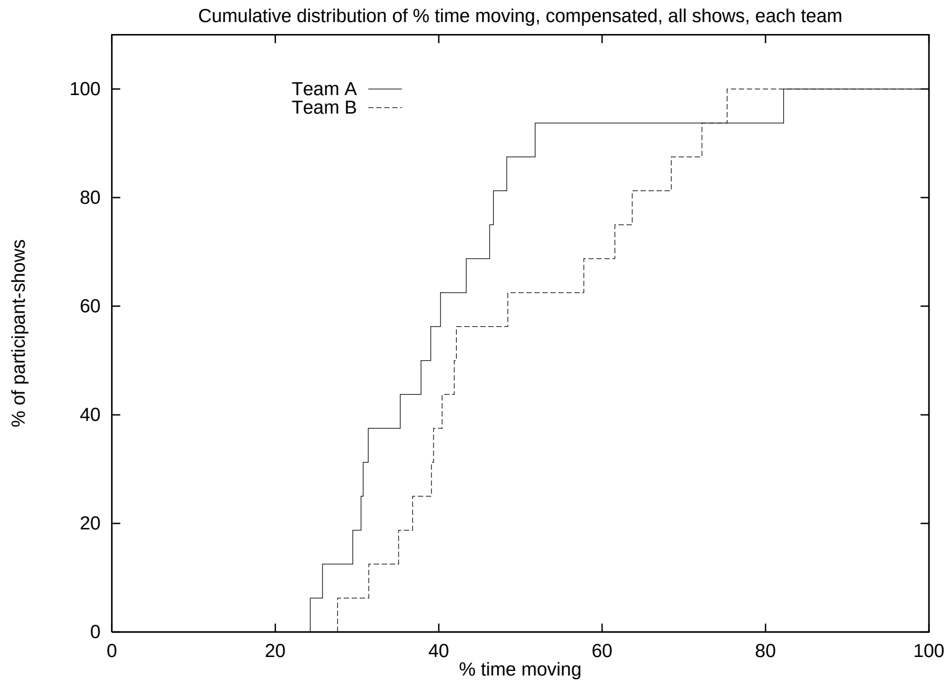

Figure 5.10: Cumulative distributions of inhabitants' movement rates, by teams.

This difference is not (quite) significant at the 90% level.

5.3.5 Movement and Shows

If we compare the movement rates for inhabitants in each show (again compensating for constraint-driven movement) we find the following per-show figures:

Show	*Mean*	*SD*	*min*	*max*
1	45.7%	15.8%	31.4%	72.2%
2	44.2%	18.7%	25.8%	82.2%
3	41.4%	11.7%	24.3%	61.5%
4	46.8%	15.8%	30.5%	75.3%

The mean is reasonably consistent for all shows. There is move difference in the variability which could easily be caused by the small number of samples (inhabitants) and outlying data points (particularly active or inactive individuals).

5.3.6 Correlation of Movement

In previous studies we have also examined the extent to which a group of participants coordinate their movements, i.e. moving independently or simultaneously. Our previous observations have been that there is slightly more concerted movement than chance, although the overall distribution is still quite close to independent movement. Figure 5.11 shows the distribution of the number of inhabitants moving simultaneously for all shows. Team captains are ignored because they move ‘all the time’, and the host is moved solely under system control.

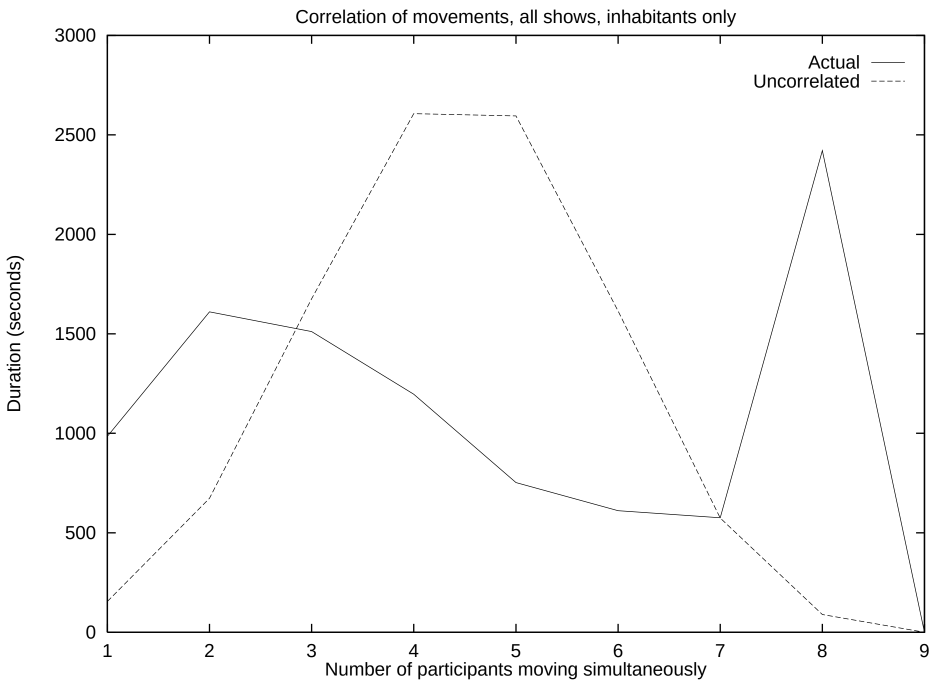

Figure 5.11: Correlation of all movements for all inhabitants.

The observed distribution is radically different from the uncorrelated curve, in particular there is (relatively) a huge amount of time when all eight inhabitants are moving simultaneously (about 10 minutes per show).

We have already noted that some movement is due to the control of the management toolkit, while some is due to the inhabitants themselves. To assess the balance of these effects Figure 5.12 shows the equivalent distribution for compensated movements, i.e. ignoring movements which are due solely to the application of constraints.

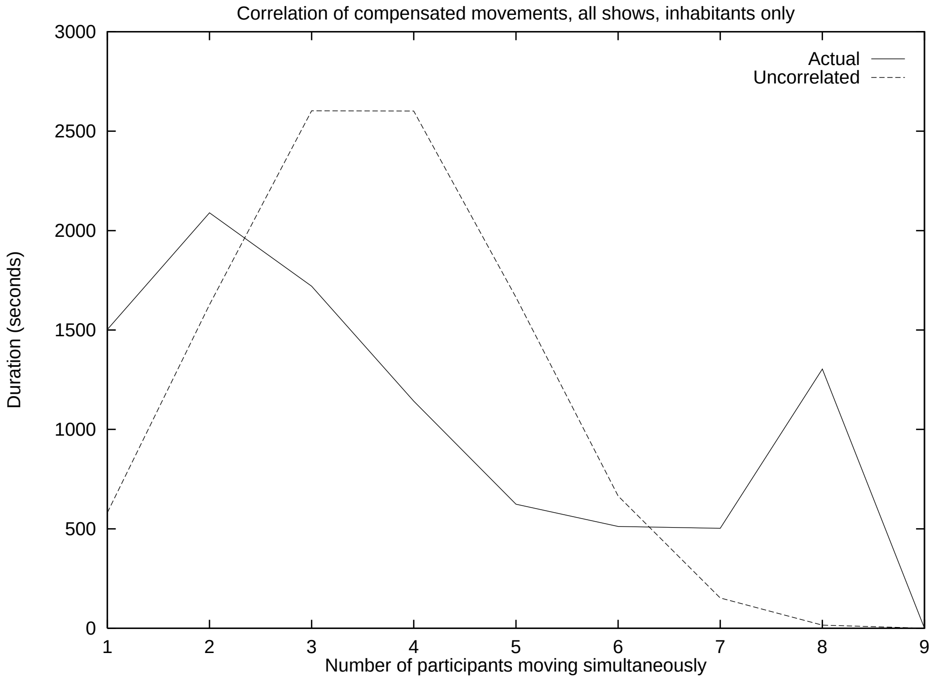

Figure 5.12: Correlation of movements for all inhabitants, ignoring constrained movements.

We note that:

- the distribution retains much of its character, with a (relatively) large peak of fully coordinated movement;
- the amount of fully coordinated movement has fallen by about 4.5 minutes per show to around 5.5 minutes per show.

This distribution is still very different from the distributions observed in previous tele-conferencing type trials. The much greater correlation of movement is almost certainly primarily due to the structure and character of the games. For example, the space frog game involves all of the inhabitants running about after space frogs, and the race involves all of the inhabitants pulling the jet car to the finishing line. Clearly, the system and network must be provisioned in anticipation of significant periods with every participant moving. This bears out the reservations which we have expressed in previous analyses concerning the likely task-dependency of measures such as this.

The additional coordinated movement is largely due the transitions between arenas on the travellators when all of the participants are moving at the same time. All of the inhabitants are also moved about at the end of each game, when they are gathered together at the exit of the arena.

5.3.7 Movement and Phases

In Section 5.2.1 we noted that there was a link between application multicast traffic and event phase structure. In this section we look at the relationship between phases and movement rates which will be one component of this relationship.

First, we consider the overall (network-visible) movement rates, due to both participant actions and management constraints. Figure 5.13 focuses on a small section of show 2, including the introductory football match and the space frogs game. It shows the average number of participants moving and the maximum number of participants moving in each five second block of the period in question.

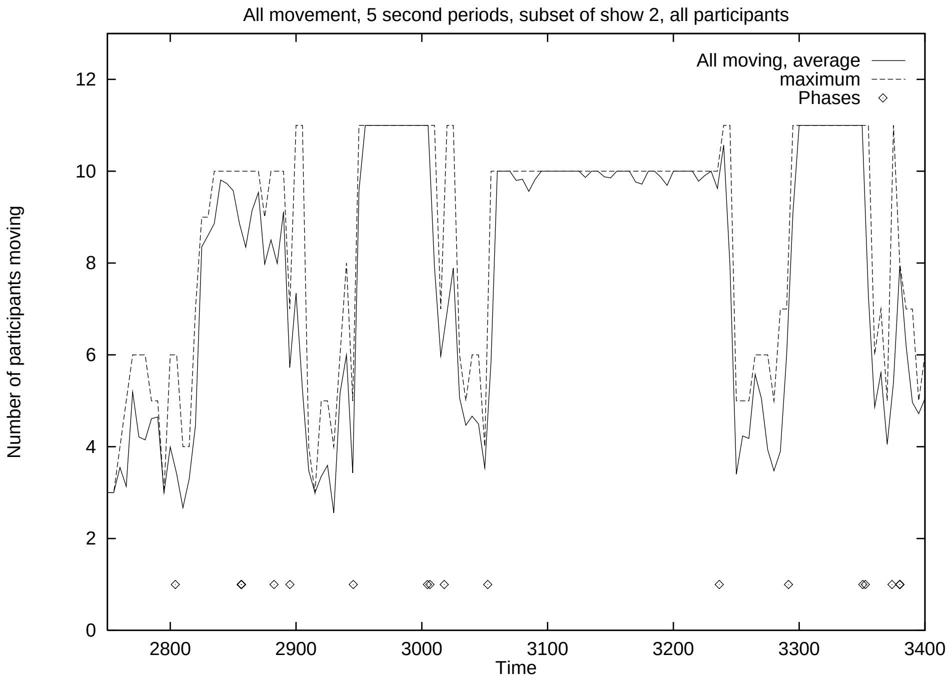

Figure 5.13: all movement activity, average and maximum number of #participants moving, for a subset of show2.

We note that:

- All 11 participants (including the host) are moving during the transitions between arenas (2945–3020 and 3290–3370) and at the start of each entry/exit phase, when the host and all participants are positioned at the entrance/exit of an arena (times 2900, 3020, 3240 and 3370).
- We already know that this view represents the team captains as moving all of the time, giving an absolute minimum of 2 participants moving, even at the quietest time. In fact, the average level never quite gets this low (e.g. during the introduction before the football match, time 2750–2820, or at the end of space frogs game, 3240–3290).

- All eight inhabitants are also moving (generally) during the football game (time 2820–2890) and the space frog game (time 3050–3240).

From the above the effect of dynamic constraints (i.e. the travellators) is clear.

If we ignore constraint-related movement and visualize the same information for the inhabitants' deliberate movements we get the view in Figure 5.14.

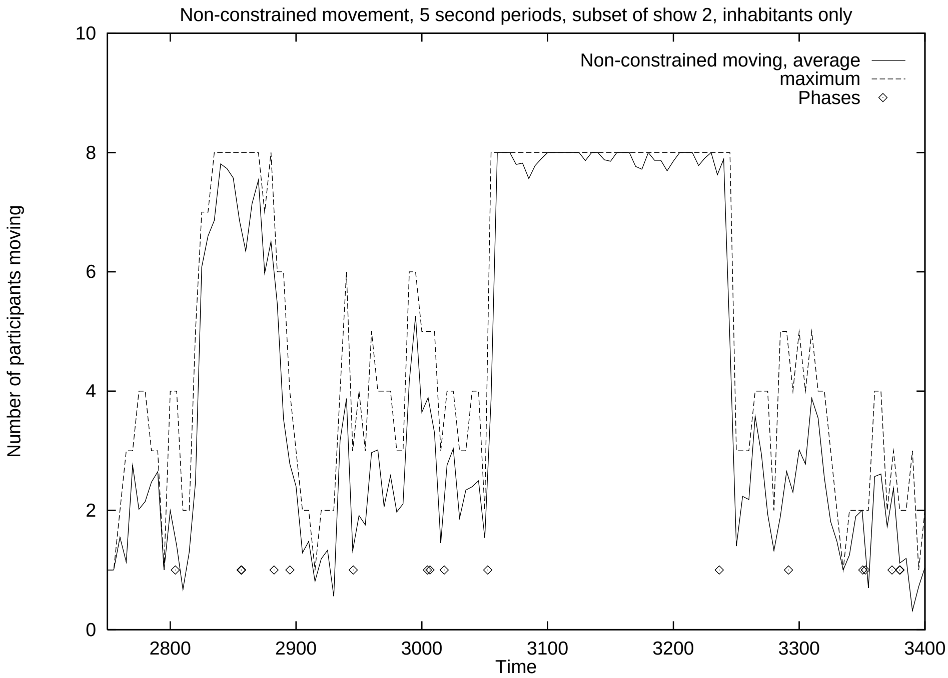

Figure 5.14: Non-constraint movement, average and maximum number of participants moving, for a subset of show2.

We see clearly that the inhabitants are almost all moving (deliberately) all of the time during the football match (2820–2890) and the space frog game (time 3050–3240), independent of any constraints. For the remainder of the period in question there is consistently less attempted movement, although the level is quite variable. Even when the inhabitants are 'pinned to the spot' (e.g. at the entrance or exit of an arena) they can still look around if they wish to. So, with the exception of the focused games, there are no clear distinctions between the different inhabitant levels of activities at other phases.

5.4 User Activity: Audio Participation Rates

Having considered participant movement in the previous section we now consider audio participation rates. Each participant (host, inhabitant and team captain) has an open microphone which they can use to talk to the other participants. In the case of the inhabitants and the team captains this is a microphone/headphones combination which allows hands-free real-time audio interaction.

5.4.1 Silence Detection

The audio communication is provided by a general-purpose audio server, one per participant, which uses packetised multicast peer-to-peer audio data distribution. In the previous trials which we have analyzed a silence detection algorithm has been used by the audio software so that audio packets are only transmitted when sounds are being received by the microphone. This creates a direct link between speaking (or at least the production of noises) and the generation of network audio traffic. It was our intention to use the same thing in *OOTW*. However, when sound-checking the system in rehearsal it was decided that the silence suppression mechanism should be disabled. This was done because the soundscape was already quite chaotic (eleven potential speakers plus sound effects) and the silence suppression was making things more difficult (e.g. losing the beginnings of quiet phrases or cutting off altogether in very quiet utterances).

Consequently, from the perspective of the network it was as if every participant were speaking all of the time! This accounts in part for the fact that audio traffic dominates the total network bandwidth, e.g. 91% of the traffic in show 2 (see Section 5.2 and Figure 5.2). In the circumstances, it is fortunate that we did not assume any lower level of audio activity in provisioning our network and systems.

Whilst the network's view of audio traffic was rather degenerate in *OOTW* we would also like to explore what would have happened if silence suppression had been retained. This will allow us to apply our experiences in *OOTW* to other settings in which silence suppression is used. It will also allows us, to a first approximation, to reason about how much participants actually spoke, and to explore whether this was influenced by the various distinctive elements of *OOTW* such as levels of participation and phase structure.

For this post-event analysis we have approximated the operation of the system's normal silence suppression algorithm by applying it to the 56 bytes of audio sample data from each audio packet which are preserved in the network traffic log (captured by tcpdump). This is 17.5% of the total audio data on which the algorithm would normally operate, and so it is likely to underestimate the amount of speaking that would have been detected had silence suppression been used in the event. Having said this, it should give quite a good approximation for our purposes.

5.4.2 Overall Audio Rates

For all participants in all shows we find that the mean level of apparent audio participation is 40.7%, standard deviation 21.8% and range 9.6% to 81.1%. This value is also higher than we have found in previous trials such as ITW (26.4%), COVEN dVS (5.2%) or COVEN DIVE (8.1%). Figure 5.15 shows the cumulative distributions of audio rates for all participants in all shows.

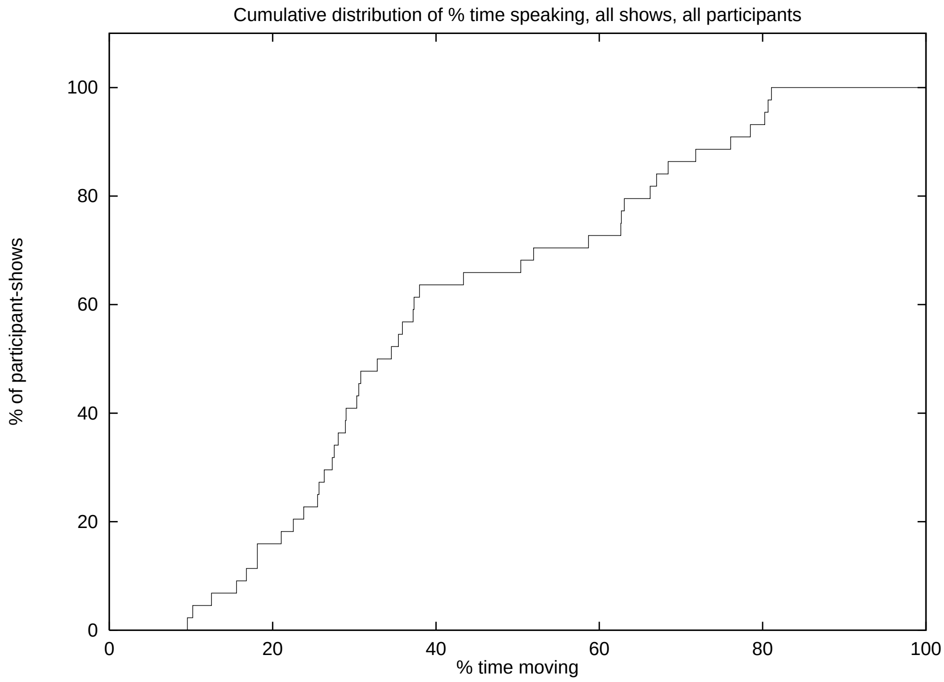

Figure 5.15: Cumulative distribution of audio participation rates for all participants.

As with movement, the distribution is clearly not normal in form.

5.4.3 Audio and Levels of Participation

Figure 5.16 shows the distribution of audio rates broken down by level of participation.

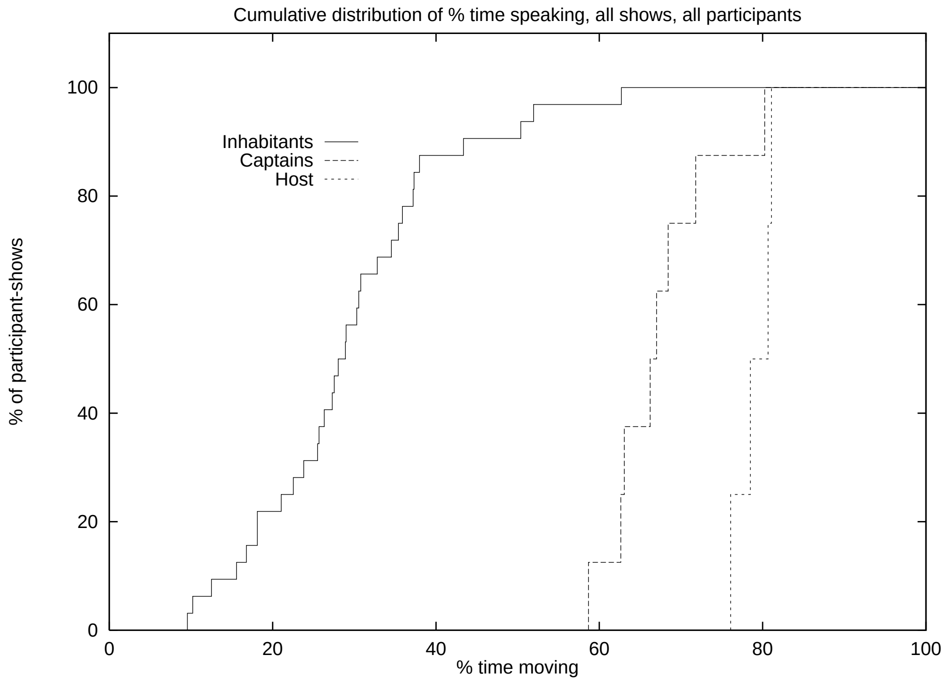

Figure 5.16: Cumulative distribution of audio rates by level of participation.

- The host has a consistently high apparent audio rate, average 79.1%, range 76.1% to 81.1%. We believe that only part of this is due to the actual speech of the host; the host was placed in the lighting booth overlooking the theatre and sound from that space (e.g. the audience and the PA) leaks back into the host's microphone.
- The team captains have a higher-than-average audio rate of mean 67.3%, standard deviation 6.6%, range 58.7% to 80.2%. This higher rate may be due to a combination of speaking more—they have clear roles and responsibilities in the event - and sound pickup from the PA—they were standing at the front of the theatre space.
- The inhabitants have a mean audio rate of 29.3%, standard deviation 11.9%, range 9.6% to 62.7%. This is directly comparable with that for the ITW trials of 26.4%, though still much higher than that found in the other trials.

We conjecture that the higher speaking rates for inhabitants may be due largely to the higher pace and more structured roles and interaction in *OOTW* compared to the previous trials in which we have been involved. These have tended to be rather pedestrian and disorganized, with no clear patterns of responsibility and less focused involvement.

5.4.4 Audio and Teams

If we now compare the audio rates for the two teams we find:

- team A (female) had a mean audio rate of 27.6%, standard deviation 11.4% and range of 9.6% to 50.4%;

- team B (male) had a mean audio rate of 31.0%, standard deviation of 12.6% and range of 15.6% to 62.7%.

The cumulative distributions are shown in Figure 5.17.

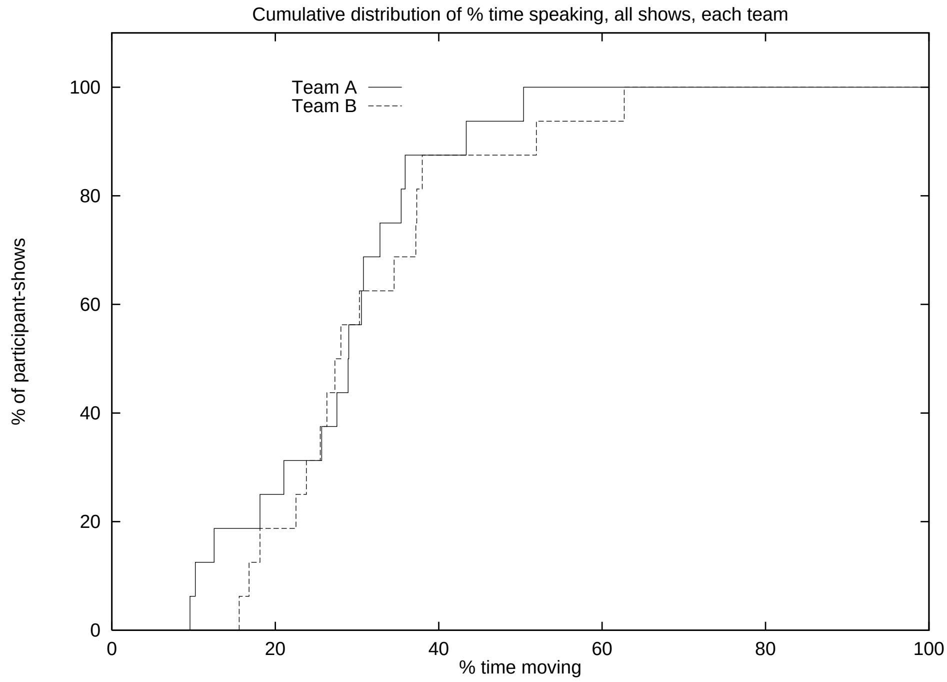

Figure 5.17: Cumulative distributions of inhabitants' audio rates, by teams.

The difference is not significant.

5.4.5 Audio and Shows

If we compare the audio rates for all inhabitants in each show we find the following per-show figures:

Show	*Mean*	*SD*	*min*	*max*
1	28.9%	9.1%	18.1%	43.4%
2	36.8%	15.5%	9.6%	62.7%
3	22.9%	9.4%	10.2%	37.2%
4	28.6%	10.4%	16.8%	51.9%

Audio rates are much more variable between shows than movement rates. However, without additional information there are no clear inferences to be drawn from this variability, other than to note its existence.

5.4.6 Correlation of Audio

As with movement we also wish to consider the extent to which a group of participants coordinate their movements, i.e. moving independently or simultaneously. Our previous observations have been that there is slightly less concerted speaking than chance, although the overall distribution is still quite close to independent movement (extremely close in the case of the COVEN DIVE analysis). Figure 5.18 shows the distribution of the number of inhabitants 'speaking' simultaneously for all shows. The values are also shown in the table below. We restrict our consideration to inhabitants because of the different characteristics of the other participants, especially the host, which might affect the outcome.

Number	*Actual*	*Chance*
1	2551	2064
2	2369	2999
3	1550	2491
4	904	1293
5	558	429
6	392	89
7	243	10
8	133	0.5

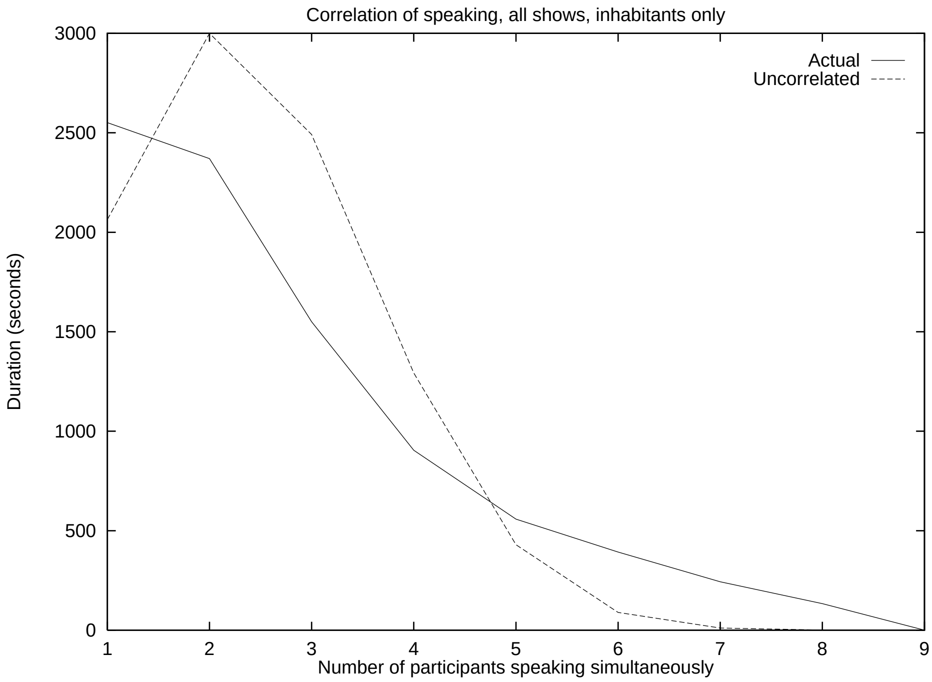

Figure 5.18: Correlation of audio for all inhabitants

The deviation from chance is much less than for movement, but still considerable. For example, one might discount the likelihood of all inhabitants speaking simultaneously based on the random model (a total of half a second over all shows). In fact, over two minutes of all eight inhabitants speaking simultaneously is observed.

We have conjectured in previous analyses that significant bursts of correlated audio activity could arise in worlds with common elements of interest and we believe that this is borne out by this analysis. Event elements such as team dynamics (e.g. cheering/heckling), response to game events (e.g. scoring, approaching the finish) and coordinated phases (e.g. being released to start a game) might all tend to produce coordinated vocal activity.

It is not possible in this analysis to rule out the possibility of external interference. For example, loud noises in the environment might be picked by all inhabitants' microphones. A much more controlled study and isolated environments would be needed to assess this possibility. However, we note that such external influences could be significant in real situations as well, for example if an on-line virtual event is linked to a television broadcast, then that would serve as a possible global source of coordinated interference. In any case, we believe that the possibility of coordinated audio activity must be considered very carefully if the quality of the experience is to be assured.

5.4.7 Audio and Phases

In Section 5.3.7 we looked briefly at the relationship between phase and movement. In this section we view (estimated) audio activity for the same portion of show 2. As in Figure 5.13, Figure 5.19 focuses on a small section of show 2 including the introductory football match and the space frogs game. It shows the average number of participants 'speaking' and the maximum number of participants speaking in each five second block of the period in question.

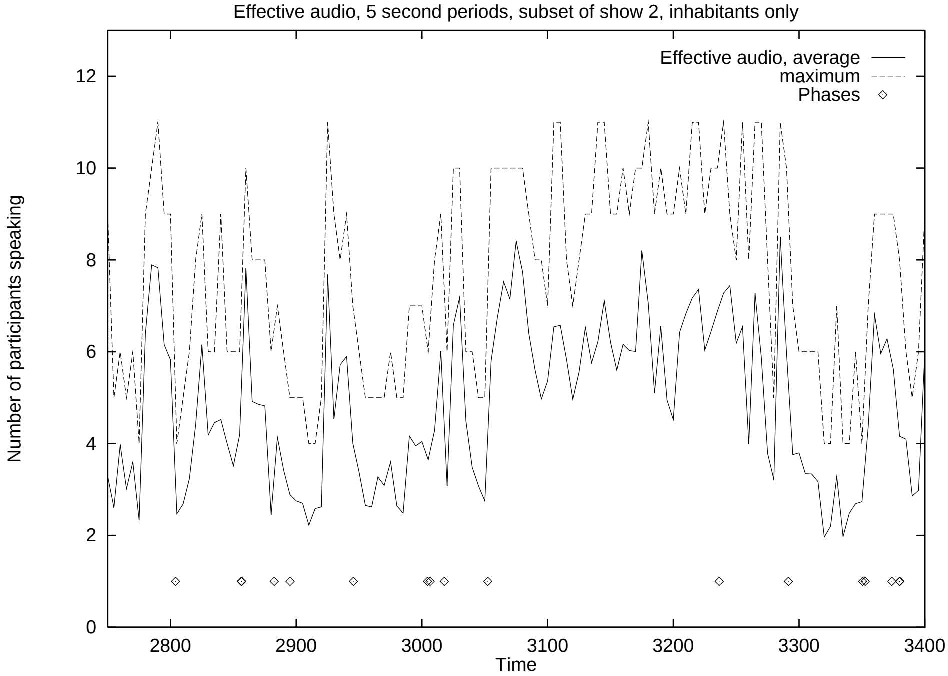

Figure 5.19: Audio activity, average and maximum number of participants "speaking", for a subset of show2.

Comparing Figure 5.19 to Figure 5.13 or 5.14 we see that the audio activity is much more variable and less clearly linked to phase structure. However, there are a number of observations which can be made:

- The largest consistent period of audio activity occurs during the space frog game (time 3050–3240) when (from observing the video we see that) everyone is shouting, giving advice, cheering, etc. There are many occasions in this period when all participants (and the host as well) are active in the audio medium.
- This audio activity persists after the end of space frog game into the post-game debriefing session; it only subsides then the participants begin the transition to arena 2 at time 3300. The participants are also relatively inactive during the transition to arena 1 (time 2945–3020), when the host is delivering a monologue introduction the space frog game.
- We can tentatively identify smaller bursts in audio activity before and after the football match (around times 2860 and 2430), on arrival at arena 1 (time 3020) and on arrival at arena 2 (around time 3360). These correspond to particularly vocally active periods in the introduction/debrief process at each arena.
- Based on only event phase it is impossible to identify times when large amounts of speaking would definitely not occur.

5.5 Summary

This completes our analysis of the network and user activity records from *Out Of This World*. We highlight the following observations in particular:

- The overwhelming majority (98.2%) of network traffic was multicast. Of this, packet audio accounted for 91%, video for 4% and the virtual world/graphics for 3.2%.
- In part, the large amount of multicast audio data was due to the disabling of silence suppression (in order to increase the quality and intelligibility of the audio mix); without this the average audio bandwidth might have been reduced by about 60%.
- There are clear differences in the network traffic generated by users at different levels of participation. In particular, the host generates the most traffic (since it includes video), the captains next (more complicated multi-part embodiments) and then the inhabitants. The cameras generate very little network traffic (around 0.5% of an inhabitant) since they cannot speak and are not visible in the world. Consequently, the division of participants between these levels can have a large impact on the overall network requirements.
- Potentially many more participants could be introduced at the level of virtual cameras, i.e. able to freely navigate the scene, selecting their own viewpoint on the event. In principle these might generate no additional network traffic. However, careful consideration would have to be given to their potential impact on the multicast reliability mechanisms.
- With regard to user movement, we find that levels of participation are again an important predictor for movement rates. Specifically, the team captains move all of the time (because of the use of magnetic position trackers) whereas the host moves only sporadically, under the control of the management interface. This is directly attributable to the form of interface chosen (or required) for each level of participation.
- The inhabitants move much more than in previous trials we have analyzed. Only part of this is due to the operation of the management interface, which moves inhabitants about at particular phases of the event (e.g. on the travellator). Compensating for this we find that inhabitants are still moving 44% of the time. We conjecture that this higher rate of movement is due to a combination of factors, including: the more active nature of the ‘show’ compared to previous tele-conferencing style applications, a simpler and more direct interface (joystick only) and the absence of alternative activities (no keyboard, no access to other applications).
- The use of movement constraints produces periods (about 5 minutes per show) when all of the participants are moving at the same time. This has potential

implications for network and client resource provisioning, since they must be able to cope with maximum levels of activity.

- Furthermore, ignoring the effects of constraints, the inhabitants still move all together a significant amount of the time (again about 5 minutes per show). This is completely different to the patterns of movement observed in the previous tele-conferencing style trials, when movement by different participants was only weakly correlated. We believe that this is due to the structured nature of the event and to the tasks which it comprises, e.g. the race requires coordinated movement from all inhabitants.
- There is a visible link between phases of the show and levels of movement. Inhabitants (voluntarily) move a great deal during the more active games. All participants (involuntarily) move during links and transitions between arenas, under the control of the management interface. At other times a greater variability of movement rates is observed.
- With regard to audio we note that silence suppression was avoided in this event for reasons of perceived audio quality and clarity of interaction. This produced higher average network bandwidth requirements (though it did not affect peak requirements, as noted below).
- By simulating the effects of silence suppression from packet data recorded in network log files we were able to explore the characteristics of audio activity in these trials.
- As with movement, we found a link between levels of participation and audio activity. The host was the most active 'speaker', although we suspect that this includes the effects of audio spill from the performance space. The team captains were active in audio much more than the inhabitants, though again there is the danger of audio spill from the house PA increasing the apparent level of speaking from the team captains.
- The inhabitants were classified as 'speaking' more than in previous analyses, although the results were quite close to those for the ITW trials. The high level in ITW trials was due, at least in part, to less effective silence detection and to the use of open speakers rather than headphones by at least one participant. We suggest that this relatively high level of audio activity (mean 29.3%) reflects the greater pace and structured interaction of this event compared to previous trials.
- Previous analyses have found very little correlation between different speakers, and in some cases negative correlation (i.e. that speakers avoid speaking over one another to some extent). However, in this analysis we found a dramatic increase in participants speaking at the same time. For example, there were two minutes during which all eight participants were considered to be simultaneously audio-active, compared with an (uncorrelated) expectation of 0.5 seconds. This casts doubt on the reliability of silence suppression as a

means of reducing peak system requirements for audio bandwidth and processing, at least for applications of this nature.

- There is some apparent linkage between show phase and audio activity, although it is more limited than that observed for movement. Most periods when all participants are speaking occur during the active games, however there are other periods of great audio activity during the more static introductory and debriefing sessions.

We return to the general character of these results in Chapter 7 where a number of ways will be noted in which they will influence our future work in Workpackage 7a's next Inhabited TV demonstrators as well as contributions to the technical context of Workpackages 4 and 5.

Chapter Six: A Field Study of *Out Of This World*

John Bowers
Royal Institute of Technology (KTH), Stockholm, Sweden

6.1 Introduction

This chapter reports on field research conducted during and around the performances of *Out Of This World* (*OOTW*). The main focus here is on documenting the production-work involved in getting the show to happen, the problems which were encountered and how they were solved. What should appear from this account is '*OOTW* in the raw'—the raw work involved in making a complex and ambitious experiment in Inhabited TV happen. The paper's fundamental issues are: Was it possible to establish practical activities within which the technology could be made to work and a show realised? If so, how, and how might the conduct of these activities (or those derived from them) be reshaped in the future either by new designs for technology or new methods of working?

It is hoped that this is of use in at least three ways. *First*, it is intended that the current paper might influence the future agendas of Inhabited TV research within eRENA, as the problems and dilemmas encountered should be more in focus as a result of this work than they might otherwise be. What requires future attention might be prioritised as a result and, indeed, this work has influenced new technologies being developed at KTH following from its results (see Chapter 7). *Second*, as this paper takes the cooperative work of Inhabited TV production as its topic, it adds to the corpus of social scientific studies of real-world settings in the literature—in particular in the literature of CSCW (Computer Supported Cooperative Work). While Inhabited TV researchers are concerned with supporting social activity, the current paper is concerned with their practical action in making Inhabited TV happen—*their* social activity. *Third*, it is hoped that discussions of 'new media', 'the future of television' and other such heady topics might be informed by a concern for these practical activities. Much of the literature on contemporary media is highly theoreticised, with speculations about new technology being more likely to be informed by psychoanalysis or poststructuralism than by study of actual work in, say, the TV or cinema industry. A recent collection such as Elsaesser and Hoffman's (1998), with some partial exceptions, only tangentially mentions what people—directors, actors, producers, cinematographers—do when making pictures amidst a main diet of cultural and film theory. The current paper, then, can also be read as a response to Manovich's (1998) call for empirical studies of the making of new media technology (here a form of interactive television) to be done now while we still have the opportunity.

6.2 Setting, Event and Technology

Let us remind the reader of the context of this research (this section can be skipped if the contents of Chapter 2 are still fresh in the reader's mind). *Out Of This World* (*OOTW*) was performed at The Green Room, Manchester on the 5th and 6th September 1998. Personnel, though, were on-site two days before and derigging was only completed on the 7th September. During this total time, the equipment was transported and rigged up, a stagger through took place along with three run throughs and four shows. *OOTW* is intended as a gameshow set in a virtual environment (a stranded, soon to implode space-station) and, as such, was intended as an experiment in Inhabited TV of a playful sort. However, as the event was performed at a venue with a performance space before an inclined area for a seated audience some compromises were necessary, as we shall see, between a staged performance and a TV simulation.

OOTW consisted of four games which were cooperative in varied ways played by two teams (‘aliens’ and ‘robots’), each consisting of a team-leader and four volunteers from the audience. All 10 players were embodied as ‘avatars’ in the virtual environment (VE). In addition, a ‘host’ was seen as a video texture projected onto a ‘screen’ within the VE. *OOTW* was presented as a TV broadcast on a large screen in front of the audience complete with title sequences and theme music. As a concession to theatricality, the (physical) team-leaders were positioned either side of screen bedecked in ‘head mounted displays’ (HMDs), some tracking equipment and pointing devices. Each team-leader had a helper by them to troubleshoot problems.

The audience themselves had one opportunity to interact with the course of events in the show through engaging with WobbleSpace (see Chapter 4) by waving coloured pieces of A4 paper. Their movements were picked up by a video camera and analysed by the WobbleSpace software for the prevalence of colours and the degree of movement of them. The same software had a role in the audience ‘warm-up’ where a version of the early arcade computer game ‘Pong’ was played with audience members using different colours to move the ‘paddles’ up and down.

At the Green Room there is an area known as ‘the mezzanine’ separate from the auditorium. It was here that the vast majority of the behind-the-scenes activity during *OOTW* took place. Sited on the mezzanine was a considerable retinue of personnel.

Four virtual camera operators captured real-time views of activity within the *OOTW* VE. A TV director made cuts from the video output of one virtual camera to another to compose the view that was transmitted to the big screen before the audience. She was aided in this by an assistant who had various responsibilities for the operation of video tape inserts into the show amongst other matters. Between them the director and assistant worked with ‘conventional’ TV direction monitors and mixing desks—the kind and number consistent with a small ‘outside broadcast’.

The eight volunteer team-members were physically located on the mezzanine together with four helpers.

One person oversaw the operation of the VR system being used (MASSIVE-2, see Greenhalgh and Benford, 1995) and another oversaw the event management application developed to work with it (see Chapter 3). Two people managed the audience interaction technology (WobbleSpace). One of the show’s two producers roved around the mezzanine area together with two researchers observing the event (one the current author, who also helped out with the music and ambient sound for the VE).

In the lighting and sound control rooms, which (unlike the mezzanine) had direct line of sight to the stage, could be found a production assistant/lighting engineer, the other of show’s producers (who also played the role of the gameshow host), a sound engineer and a musician (who played synthesised sound to give a sense of activity and spatial ambience to the games and virtual environments in *OOTW*).

6.3 Research Methods and Overview of Findings

The author's work during this period concentrated on taking notes to capture the real-time working activity of those involved in making the show. This was supplemented on occasion with video and audio recordings. Some still photographs have also been taken of the event (see Chapter 2). The style of field research conducted is broadly consistent with the 'programme' of 'ethnomethodological ethnography' practised by the author and others in a variety of settings, many involving interaction with new information technology (for a 'classic' example in CSCW, see Hughes et al., 1992). The accent in this style of research is to portray as vividly as possible the real-time details of work and social interaction as, we maintain, that it is in and through those details that coordination between persons takes place. The style of research is also descriptive, rather than theoreticised, and analytical concepts, when they are posed, are always intended to be grounded in observations made, rather than the product of theoretical deduction. An important emphasis is to capture what might be called the 'constitutively specific'—those specific features of what has been studied which make it what it is. Thus, in the current case, we are concerned with highlighting just what are the recognisable features of what went on in Manchester which made the events an instance of 'Inhabited TV' and not just an instance of some more general matter like 'an experimentation with new technology'. It is the specifics of Inhabited TV and how they were oriented to in the work of those involved which is an important topic here.

With this in mind, we now present findings under several headings: production contingencies, direction, camera work and sound, together with sections which concentrate on performer, inhabitant and audience participation in *OOTW*.

6.4 Production Contingencies

6.4.1 Budget Constraints

OOTW was not a production with an unlimited budget. Far from it. Although financial support from the ISEA festival made the event possible, this also set limits in a number of crucial respects. In particular, *OOTW* required the paid employment of a number of television professionals, most notably an experienced television director (RB). The budget set limits on how many such people and how many days could be spent in advance planning *OOTW* either at the venue or seeing the technology off-site. As one of the producers (JW) explained to the author: "The TV ideal would be for the whole crew to visit the site two times but that was not possible on the little people are being paid". For JW and RB, it would be desirable to integrate television expertise in early stages in the development of an Inhabited TV project but this has clear and dramatic resource implications for involving experienced freelance personnel on a daily rate.

While RB, for example, is very interested in Inhabited TV and concerned to influence experiments in it, there is a limit to how much time she is able to 'donate'. In addition to her involvement over the days in Manchester, RB was able to take part in an earlier planning meeting and visit Nottingham to see the Inhabited TV applications under development. However, she was not able to thoroughly explore the virtual environments that comprise *OOTW* to, for example, explore good camera angles and locations. As such she felt she was entering into *OOTW* less prepared than she would be in a real-world location shoot, where some time would be spent on site taking photographs, shooting video and thoroughly getting to know the location from the inside. Indeed, in Inhabited TV there is a sense in which there are *two* sites to get to know: the virtual environment (here the space-station) and the real environment in which the human work will take place (here the theatre itself, in other cases a studio setting).

In short, one of the most critical implications of budget constraints concerned how much advance involvement from experienced professionals could be paid for. As such person-days were limited in number, much of the 'look and feel' of the production was established with a lesser influence of television design sensibilities. Equally, the TV crew had to get themselves 'up to speed' more rapidly than might otherwise be desirable and, as we shall see, RB found herself learning how to direct *OOTW* during the course of the performances themselves.

6.4.2 Advance Readiness

The readiness of the production in advance of occupying the Manchester site was uneven. JW informed the author during the first full day of on-site preparation that he felt the system (MASSIVE-2 and its attendant applications) to be "generally well prepared but hardly tested". He felt that the system had been developed in a timely way and had not been subject to mission critical slippages. However, the integration of *all* the components and running them at performance-pace was being done for the first time on-site.

Other features of the production were similarly mixed in their readiness. There was no firm advance script or any prepared running order which could serve as a means for coordinating the work (or at least not something which was recognisable and useful to TV personnel). The absence of a script, and the involvement of volunteers, necessitated an accent on improvisation but at least one of the performers (team-leaders) informed the author that she was more "at home with scripted character development". These absences particularly concerned JW (and others familiar with television practices) who ensured that a running order was available in some form before the 'stagger through'. Some semi-scripted components were also introduced to *OOTW*, particularly for the role of the host and some prepared lines for the performers. However, this writing work was largely done during the time on-site. Interestingly, this contrasts with the advance preparation of the event management software which itself contains means for 'scripting' the activities of participants (or at least to the extent that these can be technologically constrained, see Chapter 3). That this technical sense of event management and scripting was not enough to give personnel (especially those of a television background) enough support in their work and its coordination is a topic we will return to.

We have already noted that RB, the director, had not had the opportunity to thoroughly acquaint herself with the VEs of *OOTW*. For their part, the four camera operators also had little advance practice composing shots within the environments. The exploration that they had done was confined to free navigation. They had not had the opportunity to try out, let alone thoroughly get to know, the camera control application that had been built for *OOTW*. This compounded RB's lack of working knowledge of the environments as it was through the shots given her by the cameras that she was to come to know *OOTW's* VEs.

6.4.3 Setting Up Equipment

As might be anticipated with such a 'technology-dense' event, setting up the equipment and making it all work proved problematic on a number of occasions. Early on, some machines refused to boot. Another refused to recognise its video card. The director's TX monitor, which shows the image transmitted to the large on-stage screen, failed on the first full day on-site and a substitute had to be found. The projector used to front-project the TX to the large screen did not work satisfactorily initially. A repeated pattern of interference could be seen every seven seconds which could be only eliminated by projecting a black and white image. Various ad hoc experiments were conducted to attempt to troubleshoot these problems as the producers, the Green Room's in-house engineer, the director and others gathered around. Cabling was examined for obvious physical faults. Single S-video cabling was compared with double SVHS. The image on the TX monitor was manipulated and compared with the projected image—a troublesome procedure as the TX monitor was on the mezzanine, the screen in the auditorium, and people had to run between the two to check the differences (and then the TX monitor itself failed!). The projector was moved to try and shorten cable lengths. New cabling was substituted. Finally, one of the mixer desks was removed and the interference disappeared. Accordingly, a substitute mixer had to be locally sourced at very short notice.

Naturally, such instances of creative, collective troubleshooting are as familiar as the fact of unforeseen contingencies. The important point to emphasise about Inhabited TV, though, is that such contingencies are, in a sense, raised to the second power by the coexistence of so much electrical and electronic equipment and of such varying kinds. Computer equipment and TV equipment co-exist, and in abundance. The possibilities for complex interactions between equipment is visibly apparent from the amount and length of cabling involved, something which a number of the TV crew expressed surprise and worry at.

6.4.4 Ordinary Artefacts, Everyday Methods: Especially Paper, Reading and Writing

Not all troubles were as irksome as the failing projector and mixer. Several could be addressed by more ready-to-hand and, for those who practice them, everyday methods. All graphical virtual world software was restarted by CG about 20 minutes before each performance—a routine thing to do for the sake of "memory cleanliness". Written records were persistently kept for the most critical of details—especially when computers could not be trusted with necessary information or did not provide it in a usable form. For example, the audio levels on the PCs used by the team members could be lost from one boot to another as the machines reset to default levels. So AB, one of the helpers for these participants, kept a hand written record on a scrap of paper folded in his pocket. Indeed, my field notes contain multiple instances of people keeping such notes and discussing matters in relation to them.

One should not be surprised by the obduracy of hand written paper notes in such a technology rich setting as Inhabited TV production. After all, many of these notes concern how to get the technology to work in the right way in the context—*this* context (an Inhabited TV event, here and now). None of the technologies were designed with the intention of eliminating any paper based artefact such as a 'running order' or a 'script' as those terms are conventionally understood by TV workers. This, however, is worth closer examination and clarification as Chapter 3 describes the *OOTW* event management software as supporting, amongst other things, the scripting of the event. However, the schedule of events within this software (some 67 phases with different profiles of inhabitant-control, object behaviour, default camera positions) did not constitute a 'running order' that was recognisable by TV personnel. Instead, JW spent some time typing a running order into his PowerBook which was printed out and distributed. This was not because the event management software would yield a bad 'running order' but that it didn't embody a running order at all. The detailed description of default camera positions, object behaviours and so forth is necessary to shape the behaviour of MASSIVE-2 during the show but this is a very different thing from giving producers, directors, lighting and sound engineers answers to questions such as *what happens next? what serves as my cue?* so that they can shape their own behaviour accordingly. It is answers to these questions that a running order provides and which the event management software was not designed to produce. A running order and a script speak to participating persons, the event schedule in the management interface controls technology.

Once this is clarified, the persistence of paper should not be seen as critical for any of the technology involved. It is around hand held scraps of paper (often a copy of JW's running order) that conversations were observed: "first this", "then that", "wait till you see or hear this or that notable thing". The event management interface did not exhaust the need for 'traditional' documents and other resources, especially those of TV origins. Indeed, MC, who operated the event management interface, himself had a running order ready-to-hand with annotations to help him identify cues for triggering a new phase, just as RB had a running order to guide her to select shots and cue others in turn. A basic running order that could be swiftly manually modified and moved around was what was necessary and ordinary paper artefacts with their familiar interaction methods remain just right.

6.4.5 Using the Real World Space

Though an experiment in Inhabited TV, *OOTW* was set in a theatre space and, as we have noted, attempted to simulate broadcast in a number of ways. However, TV simulation and theatre show lay uneasily alongside one another and several difficulties had to be resolved. For example, it was felt quite late in the production work leading up to the show that real performers should be incorporated into *OOTW* to satisfy the demands of theatre. The question then arose as to where physically to locate performers.

After the possibility of locating them within the audience was rejected on the grounds of potentially confusing discrepancies between world-sound and PA-mix (an issue we'll return to when discussing sound), it was decided to locate the performers on stage and 'immersively' (sporting HMDs). For JW, it was highly questionable whether this "concession" to theatricality was "crucial to the core of Inhabited TV" and perhaps for these reasons exactly how the performers were to occupy stage-space and what use they were to make of it had been underdeveloped in the lead-in to the show (an issue we'll return to when discussing the performers).

Resolving the tensions between theatre and TV simulation, then, had consequences for the use of real-world space. On a similar theme, the question of where to physically locate the TV crew and the team members raises itself. Again, if the audience had line of sight to them, the plausibility of the event as an Inhabited TV simulation might be jeopardised. Fortunately the Green Room was equipped with the mezzanine area located to the rear of the lighting and sound control rooms at a raised level above the entrance and bar. Double doors could sound proof the auditorium from the mezzanine and, again fortunately, the noise from the bar, to which the public has access independently of attending shows, was never loud enough to distract the TV crew or find its way into the mix via a team member's headset. Again in the name of TV simulation, it was seriously proposed to place physical screens or curtains to occlude team members (or at least teams) from one another and from the TV crew, as dispersed home viewers would not be able to see each other or the crew under broadcast conditions. This was still an open question Friday evening, the day before the first performance. Ultimately, the proposal was abandoned as the venue itself did not have screens available and, on the day, more pressing matters required attention.

The use of an on-stage screen also raised questions about how the theatre space was to be used. A screen would have to be sited so that (i) the front row is not confronted with an uncomfortably large image or (ii) the back row with an uncomfortably small one, (iii) an adequately large image is possible from where the projector is located, (iv) there is adequate space for on stage performers without compromising the audience's line of sight to the screen or causing stage lighting problems, (v) adequate lengths of cabling could be found to connect mixer and projector (not a trivial matter as this was a larger 'throw'—from mezzanine to auditorium—than commonplace), and that (vi) some sense of the adequacy of TV simulation could be maintained. The Green Room possessed a permanent fixed screen sited to the very rear of its stage space. On arrival JW, although he had visited the venue before, noted that the distance from the front row of seats to this screen was "larger than I remembered". The distance would be appropriate for establishing a comfortable performance space directly in front of the screen as might be required for shows with a critical emphasis on the theatricality of on-stage activity. This deeper staging was not appropriate for OOTW and would almost certainly compromise several of the features listed above. For example, it would be impossible to site the performers each to one side of the screen without withdrawing them excessively to the rear of the performance space. Siting the performers nearer the audience would almost certainly occlude the screen for many audience members as well as compromising the intention of a TV simulation. Finally, it was also determined that the in-house screen was inadequately reflective for a large, sharp image given the distance from the projector.

Accordingly, JW decided to locally hire another screen and erect it nearer the audience. Fortunately, an appropriate screen could be hired in at such short notice and a position was found, after no little experimentation, which adequately addressed (i) to (vi) above.

We have documented here a number of examples of how discrepancies between the demands of theatre and the wish to simulate a TV experience relate to specific issues about the use of real-world space and where and how things may be deployed in it. Without artful arrangements in real-world space, the event could not have been plausible as a simulation of a virtual world event, still less as a TV simulation, and let alone as a theatre experience. Other deployments under more ideal circumstances may have contributed to a better simulation or a better theatre show. The important point for an ethnography of the production work is to note the profound contingency that was experienced between the producers' and researchers' wish for an informative experiment (which might yet be adequate theatre) and how the physical space was used—a contingency which required much work attending to it, little of which was anticipated.

6.4.6 Timing

Performance times were fixed and had been agreed with the venue and organisers of the ISEA festival long in advance. As JW put it when briefing the crew on Friday morning: "We are going to perform six thirty and eight thirty on Saturday and Sunday no matter what, short of a meteor impact". While this degree of aggressive temporal constraint had been experienced by some of the academic researchers before in earlier Inhabited TV events, to those from that background new to live shows, this took a little getting used to. A relatively relaxed early Friday evening trip to a pizza restaurant was brought to a close by JW entering and complaining that the agreed time to be back had passed 15 minutes ago: "You are late, this is serious, you must learn television time." Thereafter, when addressing the crew, JW was very precise about timings and their significance. For example, early on Saturday morning: "We will run through at 12, 2 and 3.30. This means everyone in position, on cans and ready to run". RB, for her part, was commonly asking about timings, whether "we are on schedule", finding out about slippages and ensuring everyone she encountered knew any revisions.

For RB, this sense of timing comes from the production contingencies of TV. Broadcast schedules are non-negotiable and can give program start times to the minute. A delay is a very serious matter. When a delay is accountable, "heads can roll" and adjustments to published schedules require permission at "the highest levels". In commercial television, advertising breaks are often automatically scheduled with the consequence that an over-running program might simply be cut. To cope with this, RB as a director of live shows will have a personal assistant "whose sole job is to count down the minutes and seconds to the next break against the running order and tell me to speed up, slow down or keep on course". The scripting of recorded shows often involves timing scenes or even camera shots to the second so as to ensure efficient editing into the desired broadcast length (a rarer constraint in cinema). Just as learning 'TV time' needed greater precision for those from a research background, adjusting to 'theatre time' was a (slightly more welcome) feature of the work of the TV crew. RB didn't have to prompt the performers about timings during the shows. She could concentrate more on getting the right kinds of shots from the camera operators and cutting between them. A few seconds lost might have aesthetic but not legal or future employment consequences.

As it turned out, the start of the first show on Saturday was delayed as the theatre box office could note cope with the mass of people arriving at the last moment having just hurried from other ISEA events. This delay was fortuitous in some regards as we shall see. Even so, the crew went into the first show with some known problems (e.g. with the audio from one of the on stage performers, with some of the music unrehearsed) and “on a wing and a prayer”.

6.4.7 Being Aware of and Explicit About Contingency

The experimental nature of *OOTW* enabled the crew to prepare itself and to calibrate the expectations of others in ways not possible with conventional broadcast or theatre performance. *OOTW* was advertised as an experiment and JW addressed the audience for a few minutes before the title sequence rolled at the start of each show. He emphasised its experimental status and his anticipation that something would go wrong. He made the tentative yet innovatory nature of *OOTW* clear to the audience and noted that they were participating in what could be the emergence of a new medium. He invited the audience to stay in their seats after the show and discuss it with the crew, being as critical they wished—”we have thick skins”. It was also made clear before each show that the audience could leave by a rear exit onto the mezzanine and see the equipment and behind-the-scenes personnel that made the show possible. All of these are methods for conveying the contingent and experimental nature of *OOTW* to the audience.

Similarly, an earlier briefing from JW, in addition to explaining what to do in the event of fire, how their expenses were going to be covered, that they should keep the theatre tidy, advised the crew that things were expected to go wrong but that “like in a Grand Prix, we hope a crash will make things entertaining”. Not only did the crew encounter, in Garfinkel’s (1967) phrase, “the awesome contingency” of practical affairs, they were prepared for this and their audience was encouraged to be sympathetic. The contingent and experimental nature of *OOTW*, then, was public for all to see and manageable in part as a result of this. A ‘local working culture’ for *OOTW* was established for those 4-5 days in Manchester which emphasised “getting used to each other and how to communicate” and “helping out and doubling up” (JW). As RB noted to the author, this made for notable differences between *OOTW* and “normal TV” where, especially for a routine format like a gameshow which *OOTW* was emulating, people would be likely to “just know” what to do.

6.5 Direction

RB, the director, had an ‘expanded’ role over what might be customary for directing a TV gameshow. For example, she had a major task in ‘coaching’ the inexperienced camera operators, training them ‘on the job’. She artfully pushed them in run-throughs demanding a faster pace to their work than she would ask for in the shows themselves. As she explained to me: “I ran the pace deliberately quickly especially as they were inexperienced people but anyway that’s common practice so that you ease off when you’re actually going for it. That way, I and the camera operators know how much there is to spare". After several of the run throughs and the shows, RB showed the camera operators back a video and discussed it with them. Before the stagger through, an extended briefing concentrated on giving specific instructions to indicate what was required at prominent moments.

In *OOTW*, RB was combining the roles of direction (e.g. shot selection) and vision mixing (e.g. actually actioning the cut from one camera to another by switching sources at the mixer desk). This combination of tasks, commonly done by two different people, had some consequences for RB's work. For example, she was not able to experiment with anything other than the simplest of cuts—no dissolves from one shot to another were to be found in *OOTW*, or other transition effects. RB also took on a responsibility for inspecting and ensuring vision quality. For example, before the stagger through RB observed that brightness and colour were being lost and demanded that this be attended to.

Much of RB's work must be understood in terms of attempting to establish and maintain a 'working division of labour' (see Hughes et al., 1992; Martin et al., 1997) between her and other crew members (especially camera operators). For RB, this in part consisted in creating what she called 'the chain of command'. For example, she instructed the camera operators: "Even if one of the performers says something is about to happen, you wait 'till I tell you. Don't you go to the next arena merely because JW has said to the audience that we are about to". However, this had to be delicately balanced on other occasions when appropriate by allowing the camera operators their initiative and trusting their skills. During the more free-form action components of the show, "camera operators shouldn't wait for me to cue them they should just go. I can cut it off if it gets unintelligible". RB's chain of command was also moderated by the demand to "get informed on a need to know basis". When discussing the WobbleSpace software with JM, its main author, RB said: "I don't want to know other people's problems. I only want to know if something's not working. If it's vaguely not working that's OK. I needn't know." To which JM replied to RB's approval: "I'll only tell you if it's broken beyond repair".

In the terms of Hughes et al. (1992), workers commonly simultaneously maintain an *egological* orientation to the division of labour (what is there for *me* to do? what shot should I select?), and an *alteriological* orientation (what can I do to make the work of *others* easier? how can I help operator 2 get a good shot?). RB's directorial responsibilities, then, did not stop her from helping with other pieces of work—especially 'coordination work', for example, helping to ensure the cans on talkback were at the right levels or pointing at people to get them to speak when assisting the sound engineer in setting sound levels. In short, RB collaborated with everyone else in collectively managing the contingencies which arose, while nevertheless maintaining a sense of 'her job'. Equally, others maintained a sense of what they should be doing, while doing it in such a way as to make RB's job as smooth as possible.

6.6 Camera Work

Chapter 3 describes the camera control software used in *OOTW* in some detail. Essentially, it supports a number of different kinds of shot and sequence. Shots can be composed centred on various subjects (team-leaders, the centroid of the positions of each of the teams, environmental features). Preprogrammed shots and sequences can be stored and retrieved. Operators can disengage from direct control, perform multiple operations as a short sequence to define an endpoint, with smooth animation then being triggered to the endpoint as direct control re-engages. In addition, a 5DOF (five degrees of freedom of movement) ‘flying vehicle’ mode is offered for more ‘free-form’ camera work. To work with the software, RB allocated the four camera operators to different basic tasks. One operator was to follow the activities of one team, another was to follow the other. A third was to get overall views of the environment (“geography shots”). A fourth was, in many respects, given a freer rein, instructed to seek out “relationships of interest like the hand held camera would do”. Interestingly, this division of labour maps well to the different forms of camera control provided for in the design (see Chapter 3). The fourth operator, for example, would be expected to utilise the flying vehicle mode, while the team-oriented operators would be expected to use the facilities to target team-members and leaders. Finally, the operator seeking geography shots can adopt positions capitalising on interpreted hints in world definitions which allow, for example, a central location in one of the game arenas to be the object of a shot. Interestingly, RB divided up the labour of camera operation in this fashion not because she saw these as the constraints built into the camera control technology but because this would be a standard division of labour for real television analogues of *OOTW*, a division of labour which was appropriately embodied in the software.

Broadly speaking, the use of functionality in the software followed these allocated tasks. However, during the later performances as the operators’ experience in virtual camera control increased, all operators were observed using the less constrained modes (e.g. the flying vehicle) more commonly. A manually controlled shot in pursuit of a team-leader might even be preferred over automatically targeting them. Several reasons can be suggested for this. First, manual control can give the right amount of ‘camera shake’ as the target slips to one side or even momentarily out of shot. This can be more appropriate to convey a sense of frenetic activity than a shot locked to its subject. Second, manual control can enable the operator to follow the action in more flexible ways. For example, if a team member is about to come close to a frog who will then jump away (a theme of one of the games), an appealing shot is one statically targeted on the frog (rather than the team member) in which the team member looms ever closer before finally causing the frog to move, the frog’s subsequent movement being caught in a quick pan or cut. In short, it is often possible to convey action by focusing on the objects in the environment and not just the active subjects. This kind of shot was not directly supported by the camera control software (avatars but not frogs could be the centres of shots), but was commonly requested by RB, and had to be set up manually.

One should not get the impression from this that as operators become more experienced they have less need for specific support for camera control. On the contrary the kind of controlled and preprogrammed shots we have discussed remained useful. Two particular junctures were notable. First, when there were more scripted and recognisably repeatable moments in the action. For example, just after each game, the reactions of team leaders were asked for by the show’s host. Clearly, having standard methods, provided for by the event management software and the camera interface, to obtain such predictable shots is appropriate. Second, as RB said in a briefing meeting for the camera operators: “If it all goes pear-shaped, I can always say go back to your terms of reference”. Having a known responsibility for each operator and a series of standard shots associated with it enabled the operators to emerge more or less unscathed if chaos ensued. In other words, the camera controls enabled ‘escape routes’ and provided a ‘safety net’ in times of trouble which wouldn’t be available from higher DOF navigation vehicles alone.

One of RB’s “worst nightmares” was that all cameras at a specific moment would head for the same shot and that she would have “nothing to cut to”. From time to time this can happen in real television and even with experienced camera operators. A particular fear was that this might happen at the moment a director would least wish for it: if something especially remarkable at a particular location had caught the eye of all operators. In real television, the physical embodiment of the cameras and their operators militates against this to some extent. If a camera physically moves in a certain direction to get a shot, it is often clear to others what is going on. However, in *OOTW*, it was decided not to graphically represent the cameras in the game worlds for other, good reasons (e.g. to avoid distracting the participants or occluding their views). Conventional cameras also have facilities for operators to check out the views from other cameras without having to release their own current shot. Furthermore, operators typically have visual access to a TX monitor showing the transmitted shot commonly placed on the studio floor. In this way, operators can have an awareness of what each other are doing and check on the status of what’s being transmitted. This can all help find an optimal shot or angle and avoid the director’s ‘worst nightmare’. However, in *OOTW*, the only way the operators could see what others were capturing was by physically looking over to their workstations. Additionally, the operators had no TX monitor. This led to a number of occasions where RB had a sub-optimal selection of shots available to her, especially during the more free-form action components of the show—particularly the games where the trajectories and positions of participants were least predictable.

We have mentioned a number of times that RB was training the camera operators on the job and that how the camera control software was used changed across the four shows. Equally, even though she had directed an earlier experiment in Inhabited TV (*Heaven and Hell - Live* for details of this see Chapter 1), RB was also learning on the job. Specifically she was learning how to direct Inhabited TV, how it differed from conventional live television, how it was similar. Notably, she made a number of changes in how she directed *OOTW* between the Saturday and Sunday performances, making it clear that she had been “thinking long and hard over what wasn’t quite right about it”. Certainly, a number of audience members on Saturday complained that the cutting from one viewpoint to another had made it hard for them to enjoy the show.

When the crew reassembled on Sunday, RB briefed the camera operators. She played back video tapes of the Saturday performances and made numerous comments and suggestions about the camera work. She complemented the operator with responsibility for searching out relationships on the quality of his shots and encouraged the others to learn from what he had done. She urged the operators following the teams to “get more involved with movement and follow the action”. She asked everyone to “think about framing”, the internal composition of shots, and amongst other things to offset their views more so that they were not so fixed on a centre point. She suggested strategies for dealing with problems so that, not only might the operators correct themselves, they would still be giving her useful shots while they were so doing: “if you hit nothing by going left, peel back and widen while you come back, then zoom when you have something”.

For her part, RB resolved, as she explained to the author later, to make fewer cuts in the Sunday shows. Especially if the camera work was improving she could stay with individual operators longer before cutting away to another. That is, the overall show would be “cut slower” with the content of shots rather than rapid editing being used to convey the action. Additionally, she had decided to “give up on TV convention and cross the line”. Let’s examine this more closely.

A common feature of editing practice in both TV and film is to ensure that successive shots maintain a consistent spatial sense with the participants mutually oriented to one another in expectable ways. For example, in filming a dialogue between two actors (X and Y) facing each other seated on opposite sides of the table, it would be common to set up two cameras one ‘favouring’ X (i.e. taking X to be its main subject) but over the shoulder of Y, and another favouring Y over the shoulder of X. As a pair these cameras would typically be sited on the same side of the notional line between the face of X and the face of Y. Cutting from one to another will ensure a form of spatial consistency in that X (whether full face or from the back) will appear on the same side of the screen in successive shots, Y appearing on the other side. For similar reasons, cameras at a football match, say, will typically be all along the same side of the pitch, ensuring no matter how much they pan or zoom, that the flight of the ball and the direction of play of the teams will be consistent across cuts. Departures from this either seem anomalous (e.g. if a goal keeper’s kick upfield suddenly flies in the opposite direction as a cut is made to a camera sited on the opposite side of the pitch!) or experimental, trading on the wilful disruption of conventions. A cut from a camera sited one side of the action to a camera on the other is a cut which ‘crosses the line’.

RB had deliberated for a long time about whether it was possible to cross the line in virtual reality and Inhabited TV in ways which it was not in conventional television and film. While she was worried that crossing the line might seem just as jarring, RB was beginning to be tempted by the possibilities for “getting into the midst of the action”. Furthermore, for some of the games in *OOTW*, it was often hard to see where the line of action was anyway (if any). Much of the action was distributed as, for example, multiple participants chased multiple frogs. Even when a clear direction to the action was present (e.g. in the final race game), it was a common experience that the camera operators were giving her a set of shots which would often necessitate crossing the line anyway. Each of the camera viewpoints being mobile, it was not uncommon for the four shots to be taken from virtual locations all around the environment. So, sometimes there was little choice but to cross the line.

Overnight RB began to convince herself that crossing the line might be no bad thing in Inhabited TV and that she would not pass over well composed or interesting shots merely because cutting to them would disrupt a conventional sense of spatial consistency. This together with the other changes in editing style and camera practice yielded two shows free from any further audience complaint about disruptive editing (they complained about other things instead!—see Chapter 7). RB expressed it in confident terms when the author visited her the week following *OOTW*: “I think the reason why most directors are so hung up on not crossing the line is that they can’t get close to the action. But when you can get close, why not? It can be very effective. I think this is especially true in VR where there are no physical obstacles to you viewing things close to”. For RB, the realisation that the line could be crossed was a significant advance in her understanding of how to direct Inhabited TV reflecting some unique opportunities for picturing action in virtual reality.

6.7 Sound

A significant set of difficulties, confronted by RB and her camera crew in *OOTW* which often inhibited effective camera deployment, concerns the quality of the virtual world sound. As RB expressed it: “In TV sound leads vision”. It is commonly audio events which cue the director and camera crew as to what should be the subject of a shot and which shot should be selected. Someone starts to speak, a cut to them is usually expected. Being able to practically act upon the basis of sound in this way was compromised in *OOTW*.

First, the packetised audio quality was not always high. Digital distortion could occasionally be heard and the audio management system gave limited support for the dynamic grouping of sounds and real-time mixing that would be common in real television. RB again: “In TV sound levels would go up and down with cuts. We need sound balances to change in line with image changes”. Second, even if a sound could be identified as coming from a specific speaker, finding the avatar that speaker corresponded to was not always easy. Furthermore, team-member avatars were not especially discriminable, all sharing a similar basic design, and visual features upon them which signified that they were speaking (i.e. the appearance of a graphical speech bubble above their head) appeared at a delay. Finally, as both director and crew only had visual access to the worlds through what the cameras were picking up, there was no way to ‘have a quick look round’ to see who was speaking. To address these problems related to sound as best she could and to help the camera operators in turn, RB suggested that men and women be split between the two teams: “I have a major problem with audio. For example, if I am showing alien and I get robot sound. I split boys and girls so I could get the relationship from the vocal quality”. The sex differentiation, now matter how controversial for other reasons (see Chapter 7), enabled RB and her camera operators to get the team right if not the individual. In addition, performers and inhabitants were encouraged to refer to themselves and others ‘by name’ more often than might otherwise seem natural: “Auntie Astra” (the name of the alien team leader), “number one”, “number two” and so forth.

This interesting dependency of vision on sound had not been appreciated fully in advance and no particular feature of MASSIVE-2, the audio server, the event management or camera control interfaces had been explicitly designed to support the ‘integration of the senses’ in the practical work of TV direction. As RB put it: “Everyone’s got set into the visuals. This happens in TV too”. Indeed, the only people associated with the production crew with some professional musical or sound engineering experience were either invited in at the last moment (the composer who provided the soundscaping for the worlds, itself a late addition to *OOTW*) or whose attention was more devoted to other tasks (the author of the current chapter, then devoted to field observation!).

The difficulties with sound were manifest in several other ways. For example, audio level setting was highly problematic for the crew and required the skilled intervention of the venue's own resident engineer. The audio sources in *OOTW*, each with their own characteristics and relevant senses of 'too loud' or 'too soft', were very many. There was sound within the VE to be made available to inhabitants and appropriately 'overheard' by the audience. This consisted of numerous voices (team-leaders and team, the host), sound effects, the synthesised soundscapes, the sound from the VT and title sequence music. Achieving the appropriate relative source-to-destination balances for all these different kinds of sound is a considerable practical problem. The front of house sound for the audience does not need to have the same mix as the sound for the inhabitants, as the sound for each of the performers, as the sound for the production crew and so forth. The host needed to be on talkback so as to receive specific instruction from the director while no other world inhabitant needed to be. While (ideally) the front of house mix would relate to the cutting of the show as noted above, one would not want those mix fluctuations being heard by the inhabitants. In short, the highly dynamic definition and mixing of sound groups would need to be supported.

However, in *OOTW*, this could not be fully achieved. To integrate with various features of MASSIVE-2 (e.g. the display of speech bubbles) inhabitant-sound had to digitally pass through an audio-server. While this—working with the event management software—permitted the definition of audio-groups, setting levels was achieved on a per-phase (see Chapter 3) not so much on a moment-by-moment basis. On the audio server, problematic sound could not be remedied by digital sound equalisation ('EQ-ing': the adjustment of different frequency bands within a source to enhance clarity). Male voices remained 'boom-y', female voices 'thin'. In contrast, the performers' voices were taken through the house PA system and appropriately EQ-ed. The clarity of the performers markedly contrasted with the inhabitants. For SB, this was tolerable giving the audience an audible impression of the different kinds of participation in the event: "After all in a real Inhabited TV broadcast, they'd be coming in over the Internet with all kinds of poor sound". While the inhabitant sound could be excused in the name of TV simulation, this led as we have seen to critical problems for direction and camera work. Attending to sound quality and usage in VEs for Inhabited TV is not merely an aesthetic matter, it has profound consequences for the work people have to do.

6.8 Participation in *OOTW*: Inhabitants

Just as JW introduced the audience to *OOTW* and shaped their expectations before the show, so did SB (the other credited producer of the show) brief the team members (or 'inhabitants' in the preferred Inhabited TV terminology). The show was introduced as an experiment in Inhabited TV: "if this was for real in the future you'd be connected on the Internet". The division of the teams in men and women was explained. It was explained how they can move using the joystick but that at certain moments control will pass from them as they will be moved to a new location, the screen flashing red as this happens.

They were encouraged to “joggle around a bit if you are not doing anything” as moving around “makes for better TV”. It was explained that they cannot see the virtual camera operators so they should expect to. They were urged “don’t fiddle with mikes” (extraneous noise from headsets being another problem with sound) and briefly taken through the running order with an overview of each game. “We are not expecting you’ll be perfect but we’d like you to be amusing”. RB emphasised that they should be prepared to initiate talk with their team leader “not just stand around and wait to be told or be spoken to, if you don’t understand say it’s number 2 here, tell me what to do” (note too how this manner of referring to oneself would also facilitate the deployment of a camera to capture a view of number 2 and the selection of the right shot, see above).

Volunteer teams were sought before each show from arriving audience members. There were considerable risks involved in this. The audience may be too small to yield the 8 volunteers required. Too many people might refuse. It might not be possible to recruit all 8 in time and brief them for the show to start promptly. For this reason, a number of crew members and friends were asked to stand by to make up numbers if necessary. Even so, SB typically had to start his briefing without a full complement of inhabitants and could not be very detailed in his instruction. The fixed performance time created a dilemma here: the longer SB waited, the less detailed he could be. Furthermore, very little time anyway was available to the inhabitants to try out the controls and get to know each other after SB’s briefing yet before the show started—typically less than 10 minutes.

In this regard, the box office delays in the first show had a desirable side-effect once all 8 team members were assembled. Between 18.30, when SB finished his briefing, and 18.47 when RB gave the one minute signal before the start, the inhabitants were free to chat and experiment with the controls. Indeed, one of the inhabitants, who characterised himself to the author as ‘a ringer’ as he knew SB and the Nottingham team very well, spontaneously organised ‘line dancing’ amongst the avatars! The humour of this and the dexterity required served very well in ‘warming up’ the inhabitants, the volunteers contributing to the show when it started in a manner which RB, SB and several others thought to be notably stronger than other groups. (There is not space here to document, from a social scientific point of view, exactly how the inhabitants interacted with each other, how they used the controls, or other features of their real and virtual world conduct during run-time of the show itself. For several remarks on this, based on quantitative analysis, see Chapter 5.)

6.9 Participation in *OOTW*: Performers

Performing in *OOTW* raised some novel challenges for the actors who participated in the production. We have already noted how one of the performers, the leader of the aliens team, said she felt more at ease with scripted material and character development as opposed to the improvisation that *OOTW* required. Even taking this into account *OOTW* (and a fortiori many related Inhabited TV possibilities) would raise challenges for actors skilled in ‘improv’. For example, the immersive equipment was difficult to perform in due to its weight and unwieldy design. Carrying an HMD for the 45 minutes of *OOTW* was very tiring for both the male and female performer. Keeping the arms raised for long periods of time to, e.g., gesture extensively with the pointing device was also very tiring.

Indeed, keeping the hands physically distant from the performers' physical bodies was necessary as hands held close to the body disappeared as separate recognisable entities within the VE. Doubtless more fine tuned calibration of the local physical body space with the local coordinates of the performers' avatars would have helped here. However, in the rush of producing *OOTW* on site, there was not enough time to solve these problems technically. Instead, SB—working with JW and the performers—identified 'rest moments' during the show, where the actors could lower their arms and be less conscious of their body movements (real or virtual) as the focus of interest was likely lie elsewhere. This was one of the main reasons why the moments before and after each game where the host discussed the progress of each team with the team leaders were introduced. If the alien leader was being asked about how confident she felt that her team's lead would stay intact, the robot leader could physically rest.

The specificity of Inhabited TV also exaggerated the problems of gesture and how tiresome sustained body movement can be. While an on-stage physical gesture from the performers would be available there and then for the audience to see, getting its virtual correlate on screen was a somewhat hit and miss affair. The virtual gesture would have to be noticed and framed by a camera operator *and* the director would have to select it for transmission. Only then would an avatar making a virtual gesture appear on screen alongside the physical performer making the corresponding real world gesture. These dual conditions were not met often enough in rehearsal before the on-stage performer relaxed or moved onto something else. For much of the time the on-stage performers could be seen gesturing but with no obvious correlate on screen. To try and correct this for the shows, the performers were enjoined to adopt gestures which were larger, more flamboyant than they might normally think of, and could be held and repeated so as to increase the chance that they'd be caught on camera and selected for transmission. SB specifically rehearsed bigger more expansive gestures with the performers and their helpers a number of times between run throughs: "It's not just you doing it, it's getting it noticed". However, this clearly *further increased* the physical demands encountered by the performers, something already found problematic, as well as occasionally putting delicate equipment at further risk. Nevertheless, the performers persevered, carefully considering when their rest moments would be, and in other ways artfully patterning their physical exertions within the shows.

Other real-world contingencies imposed themselves on these supposedly 'immersed' performers. For example, much time was spent trying to resolve contradictions between the orientation of performers with respect to the audience, each other and the on-screen image. The theatrical staging of the show might suggest an alignment of performers' (real physical) bodies to address the audience or each other, but this mutual orientation very rarely corresponded to the orientation on-screen, depending as it does on the particularities of the orientation of the avatars and the deployment/selection of cameras. Equally, there were many moments of a performer strongly gesturing 'to the wings' while their avatar was full face-on to the audience or facing the other way. In many respects, these phenomena relate to an issue highlighted earlier: reconciling a TV simulation (complete with large screen) with theatrical staging. Later discussions with audience members suggested a common strategy for reconciling this tension: ignoring the physical presence of the performers! As a final instance of the real-world impacting on the immersed performers, it is important to note the crucial role of the two helpers. The helpers assisted the performers with putting the equipment on and taking it off. They were able, e.g. when the audio of one of them failed in performance, to mediate between the performers and the production crew. They were able to stop the performers inadvertently overstretching cables or walking into the audience. In all these respects, extra human help was required *precisely because* the performers were immersed and, being so, could not always help themselves in the face of *real-world* contingency.

6.10 Participation in *OOTW*: Audience Interaction

In *OOTW*, the theatre audience had one main opportunity during the course of the show to interact with and influence the performed events and this was through their use of WobbleSpace. A booklet of coloured sheets of paper was placed on each seat in the auditorium and by waving these audience members could ‘vote’ for their favourite member of the losing team in the gameshow—the team members’ avatars rising up coloured columns in the VE depending on how much paper-waving in their favour had been detected so far (for details Chapter 4). As briefly noted above, WobbleSpace works by measuring the variation in intensity of the various colours as they are detected in a video signal from a camera trained on the audience. Calibrating WobbleSpace was problematic requiring much work from the two personnel devoted to it, one adjusting the software, the other—at various times before the first show—running between the auditorium and the mezzanine to find out WobbleSpace’s performance and to reposition or wave coloured paper.

To get a bright enough video image the house lights had to be raised but they illuminated the auditorium dimly and unevenly. Detecting a region near the rear of audience was especially troublesome. The low house lighting contributed to the misidentification of colours and of some colours more than others. Persistently, the quality of identification was in trade-off with the level of ‘noise’ (false identifications). What was especially irksome was that this trade-off itself varied with different parts of the auditorium. For example, the exit sign at the rear caused identification false alarms but, for JM, this was hoped to be a small in its overall contribution. In the work prior to the stagger-through and between the rehearsals, the testing of WobbleSpace had to be interleaved between all the other technologies calling for controlled access to the auditorium and experimentation with its lighting conditions. Scheduling these tests, then, was itself an affair requiring practical management.

Even though WobbleSpace required a full, real audience to properly test it and this would only be available in the very first show itself, JM and his colleague kept a cool nerve. Anticipating these contingencies and suspecting that it would not be possible to configure WobbleSpace optimally in the time available, they had designed in several ‘manual overrides’. The update rate could be varied so that the effect of waving the coloured sheets could be controlled. This allowed the overall amount of time in the show taken up with WobbleSpace to be influenced. If, as happened in the first performance, WobbleSpace was slow to compute ‘a winner’, the update rate could be increased. Furthermore, at the extreme, WobbleSpace could be overridden and the interactive determination of the race up the coloured columns replaced by animated processes (a facility which did not need to be used in the shows).

Making WobbleSpace work was a complex, contingent and heterogeneous affair—not just a matter of calibrating the technology, but also of careful adjustments to camera settings and house-lighting levels, of the artful management of its functioning in use, and of ensuring that audience members behaved appropriately (as well as being enjoined to wave vigorously by the host, they were reminded not to take the sheets of paper away at the end of the show!). In many respects, the practical activities of making WobbleSpace work are emblematic of the heterogeneous problem solving (cf. Bowers, 1994) required throughout *OOTW*.

6.11 Conclusion: The Working Culture of Inhabited TV

An impression should be developing by now: for all its experimentalism, technical innovation and aspirations to bring a new medium into existence, the production of *Out Of This World* was a worldly affair—an affair very much *of this world*. It called upon people's ordinary everyday skills and expertises locally deployed in managing complex contingencies. It was constrained by such recognisable matters as finite budgets and the aches in performers' arms. Technologies for enabling mass interaction were troubled by an exit light. The fluency of virtual world inhabitants was facilitated by box office delays. A lesson in TV professionalism was occasioned by slow service in a pizza restaurant.

In many radical ways, virtual worlds and interaction within them are contingent upon work in the real-world and interaction within it (cf. Bowers, O'Brien and Pycock, 1996). If anything, the contingencies, real-world troubleshooting and ad hoc improvisation of solutions to problems required in Inhabited TV are increased in number and annoyance over a conventional TV or theatre production or a research 'demo'. When *all* the work is taken into account, virtual reality often signals an intensification of real-world engagement, not an escape from it.

Inhabited TV currently lies at the intersection of a range of working cultures and raises challenges for each of them. For academic researchers, there are challenges creating technologies which need to be not merely 'demo-ed' but performed with robustly in front of live audiences drawn from the general public. For performers, there may be specific challenges for improvisational skills, (physical and virtual) bodily deportment and cultivating a gestural vocabulary which can be caught on (virtual) camera. For TV personnel, there are professional questions to do with how to direct and produce for and in VR. We have seen *OOTW's* director come to the conclusion that some aspects of conventional TV and film practice can be revised in the face of new opportunities for picturing action in VR—maybe you *can* cross the line. For technical staff, there are yet more cables.

In the words of one of *OOTW's* producers: "There's a structure of professionalism in TV which enables teams to come together quickly and do complex things". While Inhabited TV does not yet have its own *structure* of professionalism, *local practical relations between working cultures* were seen to be established around *OOTW* testifying to its *social-practical-organisational viability* as a 'new medium'—and such viability is the minimum necessary to complement working technology if Inhabited TV is to have a future. While *OOTW's* team did complex things, they did not yet do it with the speed and fluency of conventional practice. Inhabited TV is a hybrid non-conventional technology and has a hybrid non-conventional working culture to match. This, of course, for an ethnographer can be an advantage as practical affairs get explicitly debated and laid bare for all to see. In turn, it is hoped that the reader has found this account of the lived contingencies of practice refreshing when compared with the enigmas of much contemporary cultural, media and film theory.

In the next Chapter, we gather some technical implications of the ethnographic study reported here, together with reflections made by researchers and audience members, so as to inform our future research agenda for Inhabited Television in eRENA.

Chapter Seven: Evaluating and Reflecting on *Out Of This World* – Some Conclusions and Proposals for Future Work

John Bowers
Royal Institute of Technology, Stockholm, Sweden

Steve Benford, Chris Greenhalgh and Mike Craven
Department of Computer Science, The University of Nottingham, Nottingham, UK

Graham Walker, Tim Regan and Jason Morphett
BT Laboratories, Martlesham Heath, UK

John Wyver
Illuminations Television, London, UK

7.1 Introduction

In this chapter, we return to the goals we set for *Out Of This World* (*OOTW*) in Chapter 2. These goals were in turn based on our prior experience with experiments in Inhabited TV (reviewed in Chapter 1):

- could we involve members of the public in a fast-moving TV show within a collaborative virtual environment (CVE)? In particular, could we clearly engage the inhabitants with the performers and with one another, could they keep up with the action, what would they contribute, and would they enjoy the experience?
- could we produce a coherent broadcast from the action within the CVE? In particular, would the broadcast output be recognisable as a form of TV and would it be entertaining to watch?

We next present an initial assessment of the extent to which the two goals were met by *OOTW*. As we shall see, this naturally leads to some suggestions for the future work which will contribute to the next Inhabited TV demonstrator to be reported to eRENA in Deliverable 7a.2 at the end of Year 3. This will include: attending more to the design of content of Inhabited TV shows, enhancing support for camera control and direction, refining audience interaction technologies, and capitalising on what we know about the demands Inhabited TV shows place on system and network resources. We also offer some remarks situating Inhabited TV in relation to the emerging context of digital television before closing the chapter.

Along the way, the implications of all this for the workplan of eRENA are noted. In particular, as we shall see, the experience of *OOTW* has reshaped and given detail to work in Workpackages 4, 5 and 6 while also drawing from them. Indeed, Workpackages 4 and 5, at the time of writing (month 18 of the project), have already demonstrated technologies and other results suggested for them by the experience of *OOTW*. In this sense, *OOTW*, as a project demonstrator, has already effectively fed into the technical research workpackages of the project yielding new results which can be applied in the next Inhabited TV demonstrator.

7.2 Reflection on *OOTW*

Our assessment of whether *OOTW* met its fundamental goals is based on post event discussions with the viewing audiences, feedback from the performers, inhabitants and production team and opinions from press reviews. Notes were taken during the audience discussions and these were supplemented with various personal reflections via email immediately after the event and at post-event meetings. In addition, one of us conducted an ethnographic field study, reported in full in Chapter 6, which has also influenced the reflections here. In what follows, we synthesise our reflections and, where relevant, illustrate them with quotes from audience members and inhabitants.

7.2.1 Did we produce coherent, fast-pace interaction within a CVE?

Our overall sense is that we succeeded in staging a game-show in a CVE where members of the public interacted with actors around a loosely structured script. Unlike *Heaven and Hell– Live*, the inhabitants were clearly central components of the show. The pace of the action was rapid, at least when compared to our previous experiences with CVEs. The games were mostly playable and generally recognisable in form, with the possible exception of the frogs game that confused some teams and was harder to follow as an observer.

> The frogs were too complicated. [audience member]
>
> I couldn't understand the frogs. I couldn't see what my team were doing. [inhabitant]

7.2.2 Did we produce a coherent and entertaining broadcast output?

We believe that the broadcast was coherent and recognisable as TV, again to a level that we hadn't achieved with previous experiments. Indeed, as we shall see below, viewers' reactions to the piece mostly focused on the content of the show and seemed to take it as read that this was a form of television—in that sense the technology was mostly transparent.

We attribute the difference in pace and coherence between *OOTW* and our previous experiments to a combination of the production software (Chapter 3) and the design principles described at the end of Chapter 2. In particular, the ability to constrain and move participants through a series of fine-grained phases using the management interface allowed us to push the action along and sustain the overall pace of the show, particularly when combined with the use of real-time audio among the inhabitants. The success of the event structuring notation and management interface in this respect is probably the most positive outcome of *OOTW* and signifies an important direction for the development of Inhabited TV technology. The virtual camera control interface also allowed us to produce a relatively coherent broadcast, although this was a qualified success as problems remained in capturing key moments of collaborative activity such as a dialogue between two participants or key interactions with a game object—a topic we return to in Section 7.3.

It must be noted, though, that sustaining the ‘pace’ of *OOTW* was only in part a matter of how the event notation, management and camera control interfaces had been technically designed. It is also to do with how these can be used and assembled to support the cooperative work of TV production as emphasised in Chapter 6. For example, some audience members found the pace of editing in the first two shows excessively fast:

> Cutting caused me problems of attention. The shifting point of view, the sounds, people talking. It all builds up cumulatively to make it difficult to follow.

Overnight, in response to remarks like this and her own unease, the director slowed the pace of editing for the later shows and this kind of critical comment was not heard again (for more details, see Chapter 6). From the point of view of evaluating the technologies developed for *OOTW*, this is pleasing. Not only is it possible to create a coherent and appropriately paced show, there is enough scope for skilled directors to experiment with different styles (including styles which turn out to be ‘too fast’). Pace and coherence are not mechanically determined but technically supported and creatively produced. Our technologies and the *OOTW* design principles allowed, we believe, an appropriate mix of technology and the expression of established broadcasting skills.

In contrast, although applause and laughter could be heard frequently in all performances, the content of *OOTW* attracted considerable criticism in subsequent discussion with the audience as the following paragraphs now describe.

7.2.3 Lack of empathy with the show and its characters

Several viewers commented that they did not warm to the show or feel empathy with its characters. Major contributing factors to this seem to be the lack of expressive capability of the avatars and the low quality of the audio as noted previously.

> I had problems identifying with an avatar. It’s the expressions and gestures which are missing. [audience member]

> One of the problems is identification. We miss what we’re familiar with. We need other strategies for this without texture-mapped video on faces. When they win, maybe they should show more eccentric behaviour. Something to bring them closer. [audience member]

> I was straining to hear what people were saying so I didn’t want to make a lot of noise. [audience member]

> I couldn’t identify all the time with the robots. I was ready to but the cutting prevented it. [audience member after early show, see also discussion of editing tempo above]

While this lack of empathy was generally reported, some audience members were uneasy about the use of WobbleSpace to vote for a survivor:

> I felt somewhat uncomfortable about consigning someone to oblivion. [audience member]

to which an inhabitant replied:

> I was a robot in the first show. Just to assure you I wasn’t sad when I was decimated.

With the exception of adding some gestural capability to the team leaders through the use of immersive interfaces, issues to do with creating empathic avatars were not directly addressed by *OOTW*. Furthermore, applying our game design principles may have resulted in a more sparse, albeit coherent landscape that contributed to the feeling of emptiness.

7.2.4 Lack of legend and the importance of community

A further subtle factor in this lack of empathy may have been a lack of legend. Our actress commented that her character lacked a sense of history. There was no established background to the show– why were the participants on this space station? How long had they been there? What had happened previously? This lack of a shared history made it difficult to establish an interesting dialogue between the performers and inhabitants or to improvise interesting content around the framework of the show. Our impression is that a common reaction among participants was to resort to stereotypes to fill this void, in this case based on the gender division between the teams. In one show, most notably, two of the women volunteers in the aliens team spoke throughout in high pitched pastiches of girls' voices and 'ham' acted a weak-female sterotype. Resorting to such stereotypes was a major concern with *OOTW* for some of the show's viewers.

> I thought it was sexist the way there were two sexes.

Thus, although *OOTW* did succeed in establishing engagement between the performers and inhabitants through the collaborative nature of its games, the resulting relationship wasn't especially interesting, entertaining, or, for some highly critical viewers, politically acceptable.

Future Inhabited TV should invest greater effort into developing interesting characters and narratives. This might be achieved through the more central involvement of authors, scriptwriters and producers early on in the development process. This is also a clear implication of the analysis of production contingencies in the ethnographic work presented in Chapter 6. Several of the difficulties noted there could have been alleviated by an earlier involvement from key professional persons.

However, it's important to note another possibility too. The development of interesting characters and storylines might also emerge naturally from long-term on-line communities; a strength of CVE technology. In many ways, the latter approach was successfully demonstrated in *The Mirror* (Walker, 1997), where a sense of community was established over six Inhabited TV shows spread over as many weeks with repeated and sustained participation being possible.

7.2.5 Format

Our choice of a game show was repeatedly raised as an issue in the post-show discussions. This raises the further question as to the extent to which Inhabited TV should mimic existing TV formats versus the extent to which it should introduce new formats and narrative structures.

> I had difficulties with you copying a game show as it is such an established format. [audience member]

> Why do a gameshow at all? It's something with a narrow age-range appeal. You should do something more imaginative. [audience member]

Another audience member asked:

> Did you think of something which stepped outside of TV conventions?

and once the motivation for a conventional format was explained ("if we couldn't get a highly structured form of TV right then we really would have trouble") he retorted:

> Okay so you wanted to do something conventional but you could've looked at other conventions. Pantomime conventions. Physical theatre conventions.

Clearly, for this audience member, there was something disappointing about using virtual reality technology for reproducing such "closed" (in his terms) conventions as a TV game-show:

> A paradox for a technology that promises openness.

Although we would justify the choice of the game-show for *OOTW* in terms of enabling a direct comparison with *Heaven and Hell – Live* amongst other reasons, we strongly agree with those who questioned the game-show format and existing TV formats in general. A key step for Inhabited TV is to develop alternative narrative forms that exploit its novel characteristics, especially combining on-line communities, real-time narrative and broadcast TV.

We therefore argue that *OOTW* partially addressed the issues of coherence and pace raised by earlier experiments. In particular, our production software allowed us to script and direct a framework within which the public and our actors could engage one another. However, the content of *OOTW* was more problematic and content should be a major focus of future work. We summarise with the following quote from a review in the London Times:

> At this stage Inhabited Television is merely an interesting diversion hinting at greater things. One suspects it will be some time, and several more surreal previews, before the system can generate material strong enough for television. (Times, 1998)

or as an audience member put it:

> The subject matter was simplistic but the technology was interesting.

7.3 Camera Control and other Resources for Direction

Most of our discussion of the camera control software in use (Chapters 3 and 6) indicates not faults in the software but the necessity of fitting it into an integrated cooperative working environment so that, for example, the working division of labour between camera operators and between operators and director could be more effectively supported. This requires a broadening of design-perspectives as we are not merely designing camera interfaces but cooperative work applications. In this regard:

1. It is suggested that the awareness that camera operators have about the status of their and others' work could be enhanced so that, for example, they know what others are doing and where they are in the virtual environment, and what shot currently forms TX. This may also require, in contrast to *OOTW*, giving the

cameras some embodiment within the VE. However, it would not be necessary to render those embodiments in all views. Thus, the cameras could be invisible for the inhabitants or on TX (if their presence was thought distracting) while still being present 'in the viewfinder'.

2. There are good reasons for giving camera operators and the director visual access to the virtual environment independently of the cameras as such, so that operators can have a look round before composing a shot or the director can have her own view to enable more precise instructions to be given to operators.
3. The director of *OOTW* expressed the need for more direct support for conventional TV framings of shots, e.g.: "I want to be able to say to a camera operator give me the alien leader and her team favouring the leader". It is quite possible that geometrical methods might be explored to algorithmically find such shots or, rather, preliminary framings of them, which an operator could manually refine or overturn.
4. We have also noted how camera work and direction have a practical relationship with sound. This clearly needs attention.

In other respects, the features embodied in the software mapped on well to the different roles camera crew needed to take and their different activities within those roles. Importantly, the design strategy to facilitate manual control rather than to automate shots and transitions has seemed appropriate. In particular, preferring enhanced manual control at this stage of research has allowed the director we have worked with to experiment with different ways of directing Inhabited TV. She can vary the pace of cutting, give different sets of instructions to the crew, get into the midst of the action or view it from the side, and so forth. As we have seen, she consciously experimented with different direction styles during the course of the *OOTW* shows. It seems appropriate to design software at this stage which will enable such experimentation from television personnel, permitting them to address issues about the nature of Inhabited TV from their own professional viewpoint (e.g. how should one cut between virtual cameras?) rather than mandating an answer through excessive automation in software design. Technologies in Inhabited TV should be seen as *resources for practical professional activity* not as means for automating that activity.

In *OOTW*, the director's 'interfaces' were those of conventional TV (vision mixers, monitors etc.) even if what they were displaying was unconventional (robot and alien avatars within a CVE). We have already indicated that this may need extension for fully enabling Inhabited TV by, for example, giving the director camera-independent VE access. In the hurly-burly of directing work, though, it is questionable whether a fully navigable 6DOF viewpoint would be appropriate either to a director's needs or possible to fit into all the other tasks she needs to perform. What seems more appropriate is to give the director some representation of *the activity* within the CVE, where things of visual interest might be located, so that cameras can be deployed accordingly. Such representations need not be views of the CVE of a conventional (i.e. avatar-associated) sort. For example, •andor and Jää-Aro (unpublished) have proposed 'activity maps' of CVEs, computed (as appropriate) from avatar displacements, gestural activity and text/speech input measures, showing 'where the action is'. Appropriately designed visualisations might enable at-a-glance perception which could resource direction without adding a considerable overhead. Especially for very large-scale CVEs with several hundreds of inhabitants (indeed, *Heaven and Hell—Live*, the earlier experiment in Inhabited TV, had a population of 135), such visualisations might usefully aid direction and camera work and support making action (and not just entities like avatars and environmental features) the subject of shots.

As a result of the experience of *OOTW*, we have resolved to develop new prototypes for the support of camera control and direction along these lines. We are also exploring 'sonifications' of inhabitant-activity as a step to enabling sound to serve as a direction resource. We see our work as facilitating a further paradigm shift in concepts of navigation for VR. While conventional VR systems support *avatar-centred navigation* (through the control of the position of the embodiment of the user), and while we argued in Chapter 3 for *object-centred navigation* (so that movements can be made in relationship to entities in the field of view), we are proposing *activity-oriented navigation* (so that deployment in space can be influenced by the activity within it).

We will pursue this research in eRENA Workpackages 4 and 6. Uses of real-time visualisations and sonifications of inhabitant-activity as a direction resource is now a major topic within Workpackage 4 and is so as a direct result of the involvement of partners in *OOTW*. At the time of writing, early demonstrators of the integration of these with techniques for the algorithmically optimised deployment of cameras within a virtual environment have already been completed. While supporting production and direction work is the topic of Workpackage 4, we are undertaking complementary work in Workpackage 6. There, we are investigating whether our notion of activity-oriented navigation can be applicable as an alternative to the more customary 6DOF navigation tools VR systems provide.

7.4 Audience Interaction Technology

Chapter 4 of this deliverable discussed in detail the development of WobbleSpace II, the video analysis technology that was used to capture audience activity and make it influence events in *OOTW*. As made clear in that chapter, this work capitalised on research in Workpackage 6 where interaction technologies for mixed realities are studied in depth. WobbleSpace II is a clear example of fundamental research considerations in the project being applied in one of the project's demonstrators and, as such, is a notable example of the integration of Workpackage 6 with Workpackage 7a. Indeed, we have taken the opportunity of Workpackage 7a having a deliverable at 18 months to bring forward the reporting of Workpackage 6 work otherwise not due for delivery until the end of Year 2.

Video analysis technologies as a means for enabling unencumbered audience participation will be further worked on in Workpackage 6. A development strategy has been adopted so that two complementary approaches will be investigated. The WobbleSpace technology will be generalised in the form of a video analysis server which can distribute reports of luminance levels and level-changes in user-defined regions to clients which will interpret those results in whatever way is appropriate for the application in question. In other words, the video analysis 'core' of WobbleSpace II will be extracted and coupled with a simple region definition layer. A version in which by-region analysis results are communicated to client applications in the form of MIDI-controller data is already under development and will be demonstrated in a sound-control application by the end of Year 2. To complement and contrast with this simplified approach, at the ZKM, work is underway on technologies to support analysis from multiple cameras to build a sense of three dimensional interaction, something which the approach of WobbleSpace cannot and has not been intended to achieve. By working with these two approaches, a larger range of applications and user-communities can be accessed than if one alone was adopted.

The conclusions to the discussion of WobbleSpace II in Chapter 4 of this deliverable have a feature which is worth highlighting here. A notable finding of the *OOTW* experience was that audience members expected a more persistent degree of participation in events than merely being called upon once or twice in circumscribed episodes (in *OOTW*, the *Pong* warm-up and voting for the losing survivor). If audience participation in an Inhabited TV event is to be more persistent and sustained, then this raises important questions of event design and of production-support technology. Circumscribing audience participation is one way of making it manageable in production terms. New approaches to production will be necessary if (potentially massed) audience participation is to be sustained throughout a production. These are issues we will have to consider in Workpackage 4 from a general point of view and in the work contributing to the next Inhabited TV demonstrator more specifically. We return to some aspects of this point in the next Section, 7.5.

Understanding audience behaviour and expectations are also critical social scientific topics. In the current deliverable, the inclusion of social scientific expertise in studying *OOTW* has enabled a more detailed and systematic presentation of the results of audience feedback than has been possible in Inhabited TV research in eRENA before (see Section 7.2). Reflection on prior social scientific research into audience participation also informed the requirements for WobbleSpace II at the outset of its development (see Section 4.4). The contributions of social scientists is also a feature of Deliverable D7b.1, where the experience of another 'audience' (visitors to the workshop at which the rain curtain mixed reality boundary was developed) was documented (see section 5 of that deliverable). In future work throughout the project, we intend to maintain input from this kind of study so that eRENA has a detailed documentation of what our technologies and events are like to use and to experience.

7.5 Understanding System and Network Requirements

Chapter 5 presented extensive quantitative analyses of network traffic and user-movement in *OOTW*. Some of its main findings are worth recapitulating here as they have implications for future work within eRENA especially within Workpackages 4 and 5 in addition to the next Inhabited TV demonstrator within Workpackage 7a.

A clear result is how audio accounts for the vast majority of transmitted packets. Putting this together with our ethnographic results (Chapter 6) indicating how problematic the management of audio was in *OOTW*, and how difficult it was for the director of *OOTW* to use sound as a resource for her work, we have a clear indication that the management of sound in Inhabited TV needs to be prioritised as a research issue. Research is already underway as part of Workpackage 4 into techniques for sonifying the activity of inhabitants as a resource for production and direction (see Section 7.3 above). We also hope that similar techniques can be used to provide sonic representations of collectivities in CVEs to complement earlier work in the MASSIVE-2 system on supporting crowds through 'third party objects' which regulate levels of graphical detail. These techniques might support, for example, less confusing sonic renderings of a mass of participants to be available than that based on (as currently) a mixing of all their individual audio levels. Additionally, as with the use of third party objects to mediate visual awareness, the amount and kind of audio traffic that would need to be supported should become more manageable. More sophisticated audio management should have benefits for all participants to an Inhabited TV show. Such work, when it is devoted to sonifying crowds, also is a clear point of integration with Workpackage 5, where the graphical presentation of crowds is the research issue.

Another major finding of the analyses in Chapter 5 is the systematic variation that exists in network traffic and user-movement between different participants and different phases of the show. A major research challenge is to investigate whether such results can be used predictively. That is: is it possible to identify the likely resource requirements for a participant at a particular moment during a show and optimise information distribution accordingly? If so, the production support technologies we are developing in Workpackage 4 need to be designed with this in mind. For example, *OOTW* made use of emerging results from Workpackage 4 on notations for event scripting and the experience of *OOTW* will reciprocally influence later developments of this work in turn (see also Section 7.7). It should be possible from knowing the kinds of constraints that participants are experiencing during different phases of a show (defined in some event scripting notation) to make some predictions about that phase's resource demands (e.g. unconstrained constant movement and audio is likely to lead to greater capacity and bandwidth requirements than if all participants are confined to the spot while they are addressed by a performer). In cases where several hundreds of inhabitants are envisaged, being able to estimate (even roughly) the relative resource demands of different phases and participants will be vital. Deliverable D7b.1 describes how the quality of service (QoS) experienced by different users can be systematically controlled by their activity (e.g. if an avatar with a 'video face' is approached by the user, the update rate experienced by the user might increase). QoS might also be systematically influenced by appropriately designed event notations so that the right QoS is made available to the right participant at the right time.

Work is also emerging in Workpackage 5 investigating the techniques for crowd simulation developed at EPFL for their suitability, probably in some modified form, as predictive models of collective behaviour in Inhabited TV and related applications. In this way, an Inhabited TV broadcaster could be advised, for example, on the number of network addresses that would need to be reserved at a particular moment in a show to guarantee the required information transmission and receipt between participants. The analyses in Chapter 5 of this deliverable clearly point to the utility of such work and integrating it with the event notations being developed in Workpackage 4. Even in *OOTW*, a show with just 11 participants, network traffic and user-movement showed considerable variation. In shows at a much greater level of scale, we can expect such variations to create critical infrastructural problems if left unattended. As discussed in Chapter 2, *Heaven and Hell – Live*, a show with over 100 participants, used a crude *k*-nearest neighbours model to govern information distribution leading to several experienced anomalies noted by viewers (see Chapter 1). A more ‘intelligent’ approach is clearly required and our current strategy is to design event notations (Workpackage 4) and investigate using them predictively in concert with models of collective behaviour (Workpackage 5) so that likely network requirements can be anticipated.

7.6 A Note on Digital TV

It is important that we situate our work in relationship to the emerging context of digital television, not just because the first European public digital services became available during our work on *OOTW* but also because digital television possibilities will play a major role in future ESPRIT research actions.

From the point of view of an inhabitant, our experiments in Inhabited TV to date have been experienced via a computer monitor connected either to a local area network (as with *OOTW*) or to the Web (for *Heaven and Hell—Live*). For their part, viewers watch the linear TV programme on a different screen with images and sounds delivered with video technology: either projected on a screen (as with *OOTW*) or on a television receiving a live analogue signal (for *Heaven and Hell—Live*). Digital television, however, will bring these experiences and technologies together, so that a user will be able to both view and participate in principle on the same screen using the same control technology, and might indeed be able to move seamlessly between being a viewer and a participant. However, the current (first) generation of set-top boxes for digital television do not have sufficient memory or processing power to support Inhabited TV conceived in this manner. But it is likely that systems will develop rapidly, within western Europe over perhaps a three to five year timescale, so as to facilitate this viewer-user experience and to permit widescale involvement in extensive and broadly dispersed public forms of Inhabited TV.

It is important to recognise that Inhabited TV can be employed not only for innovative forms of entertainment, as in *OOTW* and *Heaven and Hell—Live*. As a new media form it holds immense promise for novel collaborative systems, including participatory dramas and structures for collaborative learning. Whatever the programme format, it is important to keep in mind one of our claims from Chapter 4 (reiterated in Section 7.4 above). Once an audience becomes accustomed to participation in digital media, the *demand* for participation becomes strong and attempts to circumscribe or limit participation can arouse disappointment. Although facilitating audience participation has been envisaged as an important component of digital television for some time, the full implications of this for programme formats, supporting technologies, the design of control devices, and the management of network infrastructures have scarcely been recognised let alone systematically researched. From this perspective, the results of experiments in Inhabited TV can be a valuable source of requirements for digital television in general.

7.7 Conclusion and Further Future Work

Inhabited TV aims to create a new entertainment and communications medium by combining traditional TV with collaborative virtual environments (CVEs) so that the public can become on-line participants within TV shows. Both TV and CVEs stand to benefit from such an arrangement; the former through greater interactivity and access to new communities and content; and the latter through additional impetus for developing communities in the first place.

This deliverable began by summarising three early experiments, *NOWninety6*, *The Mirror* and *Heaven and Hell—Live*, that demonstrated the problems of creating a basic coherent Inhabited TV show and helped define the technical research framework for subsequent work. Problem areas included: engagement between performers and inhabitants; achieving precise and co-ordinated movement; the pace of CVEs versus broadcast TV; control of virtual cameras; and lack of expressive avatars.

We then described a fourth experiment *Out Of This World* that was conceived to address some of these problems. *OOTW* aimed to create an Inhabited TV show where interaction within the CVE and the broadcast output were both coherent and entertaining and where the show exploited a real engagement between inhabitants and performers. The key technical innovation in *OOTW* was the development of dedicated production software to support event structuring and management, and the control of virtual cameras. This was combined with a set of design guidelines for the show. We have argued that this software played a major role in enabling us to create a fast-pace and coherent Inhabited TV show for the first time. However, there were still many problems with *OOTW*, both in terms of the earlier issues that it did not address but also in terms of its content. The second major lesson from *OOTW* is that greater attention needs to be paid to creating new formats for Inhabited TV, ideally ones that combine notions of community and broadcasting. There are still also considerable opportunities for refining the successful innovations in *OOTW*.

The lessons from *OOTW* may be relevant to other areas of Information Technology research, in particular research concerned with Virtual Reality in general and networked cooperative applications of VR technology as discussed, for example, in the research field of CSCW (Computer Supported Cooperative Work). For example, the idea of scripting the temporal structure of a collaborative activity and then dynamically managing it, including constraining participants' actions, is a powerful one. On-line meetings and events of all kinds might be supported through the involvement of production teams using dedicated production software. In our recommendation of this, it should not be thought that constraining participants' actions *necessarily* involves any (ethically) objectionable loss of liberty though this complaint might suggest itself to some readers. Our experience in *OOTW* is not that inhabitants complained of being (e.g.) tied to the groundplane but that they were grateful for the simplicity and easy learnability of the interaction techniques. In short, constraints can be enabling too.

The acceptability of such constraints in *OOTW* doubtless relates to the expectancies people have of television. Participants to real TV game-shows are quite used to being told exactly where to stand or being required to use certain immovable props which ensure that they are positioned in the right place and the right time. A TV studio floor often bears many marks indicating lines performers are required to stand on at particular moments to optimise camera shot composition. In short, constraining participant's activity from time to time was acceptable from the point of view of both the production crew and for performers and inhabitants probably due to *OOTW* situating itself within a recognisable TV format in which similar constraints are familiar.

In our research in eRENA Workpackage 7b, though, a different conclusion seems to be emerging. There the artists the project is collaborating with (Blast Theory) were very sceptical about enforcing such constraints in an artistic production which emphasises more the active exploration of a virtual environment by participants. Though various software-based solutions to navigation and other problems which emerged were suggested, Blast Theory persistently preferred solutions which did not involve a constraint-laden 'software fix' and often regarded the struggle with such problems as part of what could be valuable in the experience for participants. More details of this issue can be found in Deliverable D7b.1 especially Section 5. This comparison of work across the two eRENA demonstrators has implications for work in Workpackage 4 where event design and production support technologies are to be further developed. The challenge for that Workpackage is to develop a framework for understanding production support which recognises the occasions where the imposition of constraints on participant activity are advantageous *and* those where alternative solutions need to be found (e.g. through ordinary social interaction, giving advice, encouragement, or prior practice, see Deliverable D7b.1, Section 5). There may be other possibilities too. For example, we may need to provide facilities in some electronic arenas whereby the constraints participants experience are not designed in advance by some 'external agency' such as a producer or script-writer but emerge as a by-product of participant activity or through their own 'free' decision. In terms which are beginning to emerge in Workpackage 4, we need to design techniques to support 'structuring from within' (e.g. emergence of constraint as a byproduct of participant activity) as well as 'structuring from without' (e.g. the imposition of constraint as part of the production, as in *OOTW*).

As a further instance of how our *OOTW* experience may have lessons for more general contexts, our notion of 'object-centred navigation' (here exemplified in the camera control interface) and its proposed extension to 'activity-oriented navigation' may offer a novel and generally applicable alternative to conventional 6DOF navigation in virtual worlds (something that we are currently pursuing in eRENA, see Section 7.3 above). Finally, the idea of deliberately capturing and making on-line collaborative activity visible and engaging to others might also have a broader applicability, for example in other areas of entertainment or in education and training. Though some of these applications lie beyond the specific remit of eRENA, we intend to draw attention to our work from other i3 projects for whom these applications are of central concern.

We are currently planning our next experiments in Inhabited TV which will lead to Deliverable D7a.2 at the end of Year 2. Although at an early stage, our strategy is to first establish a CVE community and then to use this as a source of inspiration, legend, characters, plots and designs for a series of broadcasts. As part of this we will concentrate on refining the basic layered participation model of Inhabited TV. We aim to provide mechanisms for feedback between layers and to enable participants to make transitions between layers (e.g. so that interesting characters can emerge from the on-line community to become core broadcast content). Given the current capabilities of our CVE platforms, this may initially exploit two distinct systems, a graphics and text CVE that can support a large community of users over the public Internet and a media-rich CVE with further extended production software to support fast-pace action for broadcasting. Future technical development will focus on merging these facilities into a single system so that a large public community can be placed alongside broadcast content with real-time feedback between the two. It will also focus on extending production software, especially scripting, directing and camera control facilities. We hope that it will then be possible to create truly innovative and engaging Inhabited TV shows.

References

[eRENA3.1] Demonstration and Evaluation of Inhabited Television, eRENA Deliverable 3.1, 1998.

[HREF1] “Questions and Answers about...Audience Participation”, http://wwwphish.net/PhishFAQ/laudience.html

[HREF2] Piro, S., “It was great when it began”, http://rockyhorror.com/partbegan.html

[HREF3] “Audience Participation creates enchanting atmosphere at opera”, http://www.cc.govt.nz/MediaReleases/98audience.html

[HREF4] “Option Finder: Product Information”, http://www.optionfinder.com/product.line/product.content.html

[HREF5] “CINEMATRIX: Interactive Entertainment Systems”, http://www.cinematrix.com/

[HREF6] “Ars Electronica: Loren & Rachel Carpenter Audience Participation”, http://www.aec.at/fest/fest94e/carp.html

Barrus, J. and Anderson, D. (1996): “Locales and Beacons.” in Proc. 1996 IEEE Virtual Reality Annual International Symposium (VRAIS’96), San Jose, 1996.

Benford, S. D., Greenhalgh, C. M. and Lloyd, D., Crowded Collaborative Environments, Proc. CHI’97, Atlanta, pp. 59–66, USA, March 1997, ACM Press.

Benford, S. D., Greenhalgh, C. M., Snowdon, D. N, and Bullock, A. N., ‘Staging a Poetry Performance in a Collaborative Virtual Environment’, Proc. *ECSCW’97*, Lancaster, UK, Kluwer. 1997.

Benford, S. D., Greenhalgh, C. M., Snowdon, D. N, and Bullock, A. N., Staging a Poetry Performance in a Collaborative Virtual Environment, Proc. ECSCW’97, Lancaster, UK, 1997, Kluwer.

Benford, S., D., Fahlén 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 Academic Publishers, pp. 109–124.

Benford, S., Greenhalgh, C., Craven, M., Walker, G., Regan, T., Morphett, J., Wyver, J. and Bowers, J. Broadcasting on-line social interaction as inhabited television. Submitted to ECSCW99.

Bilson, A., J., “Get into the Groove: Designing for Participation”, Interactions, April 1995.

Blinn, J., “Where am I? What am I looking at?”, IEEE Computer Graphics and Applications, July 1988, pp. 76-81.

Bly, S., Harrison, S., Irwin, S., ‘Media Spaces: Bringing People Together in a Video, Audio and Computing Environment’, Communications of the ACM, Vol 36, Number 1, pp.28–47, 1993.

Bowers, J. The work to make the network work. *Proc. CSCW’94*, 1994.

Bowers, J., O'Brien, J. & Pycock, J. Practically accomplishing immersion. *Proc. CSCW'96*, 1996.

Bowers, J., Pycock, J. & O'Brien, J. Talk and Embodiment in Collaborative Virtual Environments, *Proc. CHI'96*, 1996.

Capin, T.K., Pandzic, I.S., Noser, H., Magnenat Thalmann, N. and Thalmann, D. Virtual Human Representation and Communication in VLNet Networked Virtual Environments, IEEE Computer Graphics and Applications, 1997.

Chalmers, M. Information environments. In L. MacDonald & J. Vince (eds), *Interacting with virtual environments*, Chichester, 1994.

Chen, M., Mountford, J. and Sellen, A. "A Study in Interactive 3-D Rotation Using 2-D Control Devices", Computer Graphics, 22(4), August 1988, pp. 121–129.

Concise Oxford Dictionary, Oxford University Press, ISBN: CN-9129, 1996.

Damer, B., Demonstration and Guided Tours of Virtual Worlds on the Internet, CHI'97 (demonstrations), Atlanta, US, 1997.

Drucker, S., Galyean, T. and Zeltzer, D., "CINEMA: A System for Procedural Camera Movements", in Proc. 1992 Symposium on Interactive 3D Graphics, Cambridge MA: ACM Press, March 29-April 1, 1992, pp. 67-70.

Elsaesser, T. & Hoffman, K. (eds.) Cinema Futures: Cain, Abel or Cable. Amsterdam Univ. Press. 1998.

Frécon, E., Eriksson, H., Carlsson, C., 'Audio and Video Communication in Distributed Virtual Environments', Proceedings of the 5th MultiG Workshop, December, 1992.

Funkhouser, Thomas A. (1996): "Network Topologies for Scalable Multi-User Virtual Environments", in Proc. 1996 IEEE Virtual Reality Annual International Symposium (VRAIS'96), San Jose, CA, April, 1996.

Garfinkel, H. Studies in ethnomethodology. Prentice Hall. 1967.

Greenhalgh, C. and Benford, S., MASSIVE: A Virtual Reality System for Tele-conferencing, ACM TOCHI, 2 (3), pp. 239-261, ACM Press, 1995.

Greenhalgh, C., Benford, S., 'A Multicast Network Architecture for Large Scale Collaborative Virtual Environments', Multimedia Applications and Services and Techniques – ECMAST '97, Proc. 2nd European Conference, Serge Fdida and Michele Morganti (eds.), Milan, Italy, May 21-23 1997, pp 113–128, Springer.

Han, J., Smith, B., 'CuSeeMeeVR Immersive Desktop Teleconferencing', ACM Multimedia '96, Boston, MA, USA, ACM 0-89791-871-1, pp.199–207, 1996.

Hughes, J., Randall, D. & Shapiro, D. Faltering from ethnography to design. *Proc. CSCW'92*, 1992.

Lea, R., Honda, Y. And Matsuda, K., Virtual Society: Collaboration in 3D Spaces on the Internet, Computer Supported Cooperative Work: The Journal of Collaborative Computing, 6, 227–250, 1997, Kluwer.

Macedonia, M. R., Zyda, M. J., Pratt, D. R., Brutzman, D. P., and Barham, P. T. (1995): “Exploiting Reality with Multicast Groups: A Network Architecture for Large-scale Virtual Environments”, Proc. VRAIS’95, 11-15 March, 1995, RTP, North Carolina.

Mackinlay, J., Card, S. and Robertson, G., “Rapid Controlled Movement Through a Virtual 3D Workspace”, Computer Graphics, 24(4), August 1990, pp. 171–176.

Manovich, L. Towards an archaeology of the computer screen & To lie and to act: Cinema and telepresence. In Elsaesser, T. & Hoffman, K. (eds.) Cinema Futures: Cain, Abel or Cable. Amsterdam Univ. Press. 1998.

Martin, D., Bowers, J. & Wastell, D. The interactional affordances of technology: an ethnography of human-computer interaction in an ambulance control centre. *Proc. HCI’97*, 1997.

Morphett, J., “eTV: A Mixed Reality Interface onto Inhabited TV”, ESPRIT project #25379 (eRENA) deliverable D3.2, 1998.

Pitas, I., “Digital Image Processing Algorithms”, Prentice Hall, ISBN: 0-13-145814-0, 1993.

Preece, J., “Human-Computer Interaction”, Addison Wesley, ISBN: 0-201-62769, pp. 281-282, 1994.

Pycock, J. and Bowers, J. Getting others to get it right. *Proc. CSCW’96*, 1996.

Reynard, G. T., Benford, S. D., Greenhalgh, C. M., “Awareness Driven Video Quality of Service in Collaborative Virtual Environments”, Proc. ACM CHI’98, Los Angeles, April 1998, ACM.

Seligmann, D. Duncan and Feiner, S., “Automated Generation of Intent-Based 3D Illustrations”, in Proc. SIGGRAPH’91, Computer Graphics, 25(4), July 1991.

Suzuki, G., ‘InterSpace: Toward Networked Reality of Cyberspace’, Proc. Imagina ‘95, pp. 26–32, 1995.

Times (1998) TV from another planet: something virtually different, The Times (Interface section), October 7th 1998.

Vernon, D., “Machine Vision: automated visual inspection and robot vision”, Prentice Hall, ISBN: 0-13-543398-3, 1991.

Walker, G. R. (1997), ‘The Mirror - reflections on Inhabited TV’, *British. Telecommunications. Engineering Journal.*, 16 (1), pp. 29–38, 1997

Ware, C. and Osborne, S., “Exploration and Virtual Camera Control on Virtual Three Dimensional Environments”, Proceedings of the 1990 Symposium on Interactive 3D Graphics, Special Issue of Computer Graphics, Vol. 24, pp. 173–183, 1990.