Broadcasting On-Line Social Interaction as Inhabited Television

Steve Benford, Chris Greenhalgh and Mike Craven
The University of Nottingham, Nottingham, UK
{sdb, cmg, mpc}@cs.nott.ac.uk

Graham Walker, Tim Regan and Jason Morphett
BT Laboratories, Martlesham Heath, UK
{graham.walker, tim.regan, jason.morphett}@bt-sys.bt.co.uk

John Wyver
Illuminations Television, London, UK
john@illumin.co.uk

John Bowers
Royal Institute of Technology (KTH), Stockholm, Sweden
bowers@nada.kth.se

Abstract: Inhabited TV combines collaborative virtual environments (CVEs) with broadcast TV so that on-line audiences can participate in TV shows within shared virtual worlds. Three early experiments with inhabited TV raised fundamental questions concerning the extent to which it is possible to establish fast-pace social interaction within a CVE and produce a coherent and engaging broadcast of this action. This paper presents a fourth more recent experiment, Out of This World that directly addressed these questions. We describe how the formulation of inhabited TV design principles, combined with the use of dedicated production software, for constraining and directing participants' actions and for controlling virtual cameras, enabled us to create a fast-moving and coherent show. We conclude that our experiments to date demonstrate the technical feasibility of inhabited TV, but that greater attention needs to be paid to developing appropriate formats and content for this new medium before it becomes truly engaging. We also suggest that our real time production and camera software may be useful in other areas of CSCW.

Introduction

Inhabited TV combines 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 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. 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 is motivated by the belief that TV and CVEs can benefit from one another. On the one hand, TV may benefit from access to the on-line communities that are enabled by CVEs as these may provide new forms of content, an important commercial issue with the arrival of digital TV. Furthermore, inhabited TV extends interactive TV to include social interaction among participants, new forms of control over narrative structure (e.g. navigation within a virtual world) and interaction with content (e.g. direct manipulation of props). On the other hand, TV may be a powerful driving force for the commercial uptake of CVEs and the development of richer on-line experiences than are provided by current chat environments.

Inhabited TV also raises new issues for CSCW in general. First, it focuses on the production and management of fast-pace on-line social interaction that follows a pre-scripted structure. Second, it raises the question of how social interaction can be captured and broadcast in a way that external viewers find engaging.

The following section introduces the idea of layered participation as a mechanism for describing inhabited TV applications. We then discuss the key challenges that were raised by three early experiments in inhabited TV, namely the difficulty of producing fast-pace and coherent social interaction within a CVE and of producing a coherent and recognisable broadcast of this action. Following this we describe a fourth experiment called *Out of This World* (*OOTW*) that addressed these challenges. We focus on the use of dedicated production software to configure the temporal structure of the show, to constrain participants' actions and to support real-time control of virtual cameras. Finally, we reflect on the experience of *OOTW,* drawing on responses from its audiences.

Layered participation in inhabited TV

Inhabited TV can be described in terms of four layers of participation. Each layer corresponds to a distinct responsibility within the show and to a distinct combination of interface and network technologies to access the shared virtual world and its contents.

Performers – the innermost layer, 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.

Inhabitants – the next layer describes the inhabitants, on-line members of the public who navigate through the virtual world, interact with its contents, are represented by avatars 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 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.

Viewers – the third 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 be socially active via off-line feedback and discussion mechanisms.

Producers – the 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. Examples include, directors, operators of the virtual cameras that capture views of the virtual world 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.

Early experiences with inhabited television

Between 1996 and 1997 we were involved in staging three public experiments with inhabited TV in order to test its feasibility and to frame key research issues.

The NOWninety6 poetry performance was staged using the MASSIVE-2 CVE (Benford, Greenhalgh, Snowdon and Bullock, 1997). The performers (poets) and viewers were co-located in a physical theatre so that each performer appeared

simultaneously on a physical stage and on a corresponding virtual stage in the virtual world. A broadcast stream was mixed down in real time from the viewpoint of a virtual camera-operator, an example of a producer. 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. These inhabitants could move about, experience the poetry and could communicate with one another using real-time audio. The event lasted for one evening and approximately 200 people were in attendance. Sixty experienced the virtual world in cycles of 10 at a time, with the remainder watching the broadcast in the theatre.

The Mirror in the first quarter of 1997 involved public access to a series of six virtual worlds on the Internet (Walker, 1997; McGrath, 1998). 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 of the virtual worlds. Edited highlights from the virtual worlds were shown on subsequent TV shows. The virtual worlds included interactive collaborative games and tricks (such as a virtual rocket launcher and bouncy castle) and an art gallery where inhabitants could display their own VRML 2.0 creations. 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) and an end of the world party. The inhabitants accessed *The Mirror* from their homes or workplaces over the Internet using Sony's *Community Place* software (Lea, 1997). This allowed for text and graphical communication between inhabitants and performers. Over 2300 people registered to become citizens of *The Mirror* and received a CD-ROM containing the browser software and VRML2.0 content.

Heaven and 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 broadcasters to viewers happened simultaneously, with the latter seeing the activities of the former. The performers consisted of a host and two contestants (celebrities on UK TV) as well as three 'reporters' who provided additional commentary. The performers were physically located in an inhabited TV studio along with the production team that was responsible for creating the live TV broadcast. This studio combined a local network of PCs, a TV outside broadcast unit and an Internet connection. The production team included a 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. As with *The Mirror*, the inhabitants accessed the shared virtual world from their using the *Community Place* software. They could also be viewers if they had a TV set in the room alongside their PC.

Games within *Heaven and Hell– Live!* included a participatory treasure hunt, a quiz and a gambling game. The intention was that the inhabitants would assist or

hinder the performers in these games. Communication within the virtual world (i.e. among performers and inhabitants) was via text and graphics. In addition, a live audio stream containing spoken communication between the performers was added to the final TV broadcast. The program was broadcast in August 1997 in a late-night slot. The on-line audience of inhabitants peaked at 135. The viewing audience was estimated at 200,000.

Lessons from these early experiments

The lessons learned from these early experiments 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. These lessons were distilled from the opinions of the inhabitants and viewers themselves as voiced over email for *The Mirror* and *Heaven and Hell – Live!* and at a post-performance debate for the *NOWninety6 Poetry Performance*.

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. However, our early experiences suggest that it is difficult to engage the public in a coherent, real-time and fast-pace narrative within a CVE.

It was difficult to establish a productive engagement between the inhabitants and the performers. At one extreme, the inhabitants were unable to get a word in edgeways as the performers dominated the interaction (Heaven and Hell – Live!). At the other, the inhabitants spent all of their time chatting to one another and ignored the performers (NOWninety6).

It was difficult for inhabitants to achieve precise and co-ordinated movement within a CVE. A conventional studio-based TV show often requires participants to move to precise locations at particular instants (e.g. standing on a mark so as to be in camera shot) and for several participants to move in a co-ordinated way.

The difference in pace between CVEs and broadcast TV was problematic. Conventional live TV shows are scripted, highly structured and have a fast pace that involves precise (i.e. to the second) timings for events. In contrast, current CVEs have a much slower pace, due to delays in interaction (e.g., navigation) and network delays. This is especially true where text communication is used.

Producing a coherent broadcast output from a CVE – the second general issue raised by our early experiments concerned the quality of the broadcast output that was created from the action within the CVE. This was especially evident in *Heaven and Hell – Live!* due to the rigorous demands of producing an hour-long live broadcast on national TV. They key problem here was with camera control. 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. In contrast, the development of CVE interfaces has focused on

controlling an individual participant's view of the world, but has not considered how social action can best be captured and displayed to external viewers.

In summary, early experiments in inhabited TV demonstrated the difficulty of establishing fast-pace, structured interaction between performers and inhabitants within a CVE and also of producing a coherent and entertaining broadcast mix of this action for external viewers. We next present a fourth more recent experiment that was intended to address these problems.

A fourth experiment– Out of this World

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 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 above. 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 gameshow. This choice allowed a direct comparison to be made between the two experiments. Given the observations above, 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, and would they enjoy the experience?
- could we produce a coherent broadcast from the CVE?. Would the broadcast be recognisable as a form of TV and would it be entertaining?

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

Layered participation in OOTW

We begin by describing how our layers of participation were realised in *OOTW*.

The inhabitants

OOTW adopted a 'cheesy' outer space theme. The inhabitants were divided into two teams, aliens and robots, who had o 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 below. 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. They were located behind the scenes, out of sight of the viewers in the theatre.

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 (left of figure 1), 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.

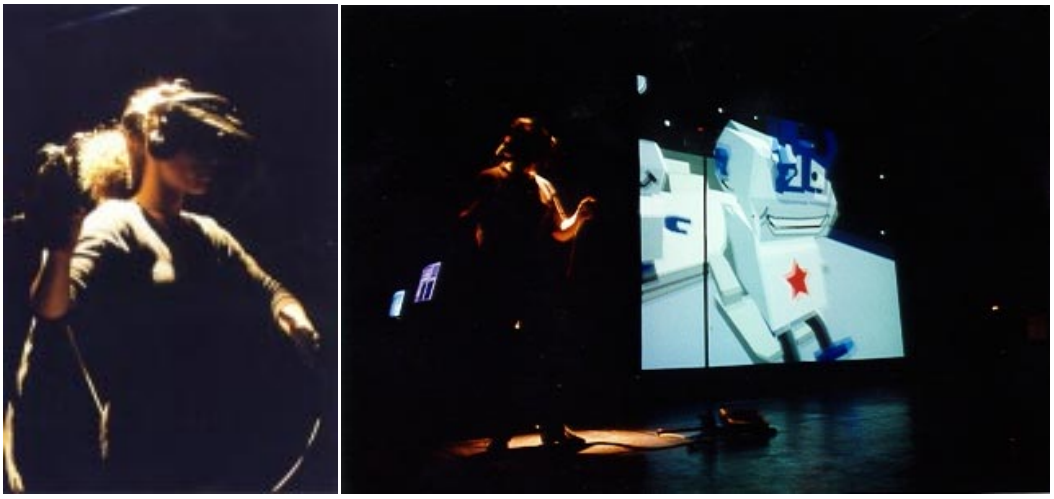

Figure 1: performer equipment (left) and location in the theatre (right)

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 (right of figure 1). 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.

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 could rotate around to show the game scoreboard.

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, that was inspired by the CINEMATRIX interactive entertainment system (CINEMATRIX, 1998). 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 automatically detected from a video image of the 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 a previous CINEMATRIX demonstration at SIGGRAPH. 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.

The production team

OOTW involved an invisible but essential production team who were responsible for managing the CVE software and for producing the broadcast. Four virtual camera-people were present in the world, although they were not visibly embodied. Using purpose built interfaces (see below), 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 below). The left of figure 2 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). The right of figure 2 shows the four virtual camera operators at their machines (physically located just behind the director).

Figure 2: The director and her assistant (left) and the camera operators (right)

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 3).

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 4).

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. 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 o 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 5).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. 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 3: the alien team

Figure 4: the alien captain

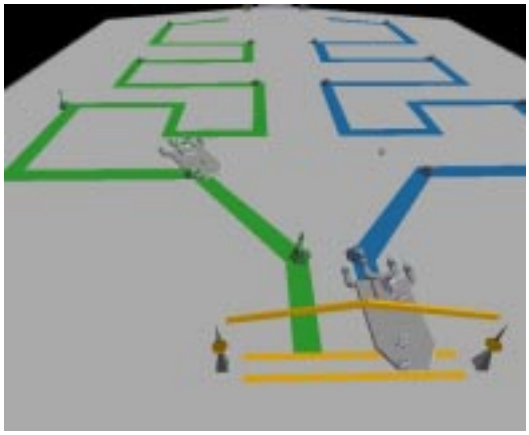

Figure 5: racing space cars

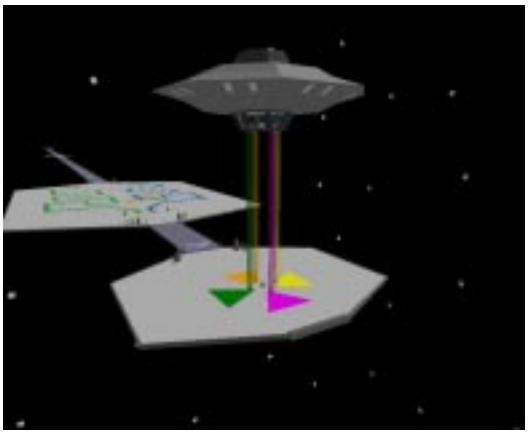

Figure 6 the space station

Design principles and production software 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.

Show design principles

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.

Production software

We developed dedicated production software to support *OOTW*. This consisted of an event structuring notation, participant constraints, a real-time management interface and a virtual camera interface.

Event structuring notation, participant constraints and management interface

All CVE platforms allow designers to specify the spatial structure of a shared virtual environment, at least in terms of its geometry and in more advanced systems, in terms of higher-level structures such as regions, locales and third party objects. In order to support inhabited TV applications, we extended the MASSIVE-2 platform to also support the definition of the temporal structure of an event in a CVE in terms of a series of phases. For each phase, the designer could configure a number of properties including:

- a name for the phase.
- objects that would be active during the phase.
- trajectories for these objects during the phase.
- hierarchical groupings of objects so that they could be attached to one another during this phase. For example, attaching a 'costume' to one of the participants.
- audio levels and extents (as defined by audio nimbus inMASSIVE-2) for specific participants.
- whether this was a roll-over phase, in which case the next phase would automatically be activated after a specified time interval.
- default positions for the virtual cameras.
- constraints on participants' movements. Each participant could be placed inside a bounding box outside of which they could not move during this phase. This box might be small enough that they could only turn on the spot or could be large enough for them to be able to explore a large area. The bounding box could have its own animated trajectory enabling participants to be gradually pulled along to a new destination during the phase, over a specified time period.

The phases and their properties were specified in a configuration file that was then loaded into MASSIVE-2. The structure of *OOTW* consisted of over fifty phases which spanned movement on the travellators, dialogue at the exit and entry points to the arenas and the structure of the games themselves. These fifty phases occurred in a forty-five minute show. Examples of the use of phase properties 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 activating objects such as the spinning rings in the quiz.

A dedicated world management interface was developed to support a member of the production team in dynamically triggering different phases as the show progressed. The phases were presented as a list and any phase could be selected

by name, causing the whole show to jump to that phase. By following a script (and taking cues from the gameshow host), the world 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 participants could be brought together into a structured arrangement at key moments before being released again into a more exploratory activity.

This ability to dynamically apply constraints to participants was therefore intended to support the orchestration of co-ordinated movements and crowd control (avoiding inhabitants becoming lost or running away) and to increase the pace of the show by enforcing time limits and by rapidly shuttling participants to new locations. In addition, the use of constrained positioning meant that potential camera shots could be established in advance and that directors and camera operators could plan for them accordingly (see below).

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. Finally, several contingency phases were specified in the expectation that participant's equipment might fail. For example, there were alternative 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.

Virtual camera control interface

The second component of our production software supported the control of virtual cameras. In *Heaven and Hell –Live!*, the virtual cameras had used standard participant navigation controls to move through the virtual space. As noted above, this led to difficulties with following the action, getting lost and having a camera's view obscured by passing inhabitants. For *OOTW*, we developed a new virtual camera control interface that was dedicated to the task of capturing the action in a CVE from a third party perspective. At the heart of our approach was the idea of object centred navigation that we had first tried at the *NOW'ninety6* poetry performance. This involves locking a virtual camera onto a specific focus or target and then controlling it in such a way that the target is not lost from view and can be framed appropriately, for example, zooming in or pulling back to show its relationships to other targets. Our design was also intended to facilitate artistically engaging camera work, involving the kinds of long sweeps, zooms and tracking shots that can be seen in expensive movies and computer animations.

OOTW introduced a wide range of potential targets including scenery, individual participants and the teams. We addressed three major considerations in designing a virtual camera interface to cope with this level of complexity:

Target selection – we provided three ways of specifying the current target of a virtual camera. First, the operator could jump to pre-set fixed points in the virtual environment. These were selected from a list in the camera interface and

included key locations, defined separately for each phase of the show at configuration time, as well as locations that had been previously marked by the operator. Second, the operator could choose to track a single participant or a group of participants (e.g. one of the teams), again by selecting them from a list in the camera interface. The camera would then dynamically adjust its position to follow the target as it moved. For groups, it would take the average position of the group's members. Third, more conventional free form flying was also provided of the type that would normally be associated with an inhabitant interface.

Relative viewing control – we then enabled the camera operator to move the camera relative to the target. The operator could use independent sliders in the interface to control the yaw, elevation, distance and vertical offset of the camera relative to the target (the position of the camera in relation to the target was described using spherical polar co-ordinates). These controls allowed the camera operator significant control over the framing of the target within the shot.

Temporal control – although the target was selected directly and interactively, the relative viewing controls could be applied with three different timings. With real-time control the camera would move as each slider was moved. This was subject to a controllable damping coefficient so that the operator could trade off responsiveness for speed of movement. With just-in-time control the operator could disconnect the sliders from the camera, use several different sliders to adjust different relative viewing parameters and then apply the changes as a single atomic operation. With pre-programmed control, the operator could define and store sets of viewpoint parameters. Selecting a set of parameters would trigger a smoothly interpolated movement to the specified position. In addition, in both the just-in-time and pre-programmed modes, the operator could build up a sequence of camera moves to be triggered one after the other and then step through this sequence using a single interface button.

As noted above, *OOTW* employed four virtual cameras whose operators were given different roles by the director (e.g. following different participants or capturing a birds-eye view). The camera operators and director were also on a live talk-back system so that they could communicate freely during the show.

Reflection on OOTW

We now present an initial assessment of the extent to which the two goals described above (involving the public in a fast-moving enjoyable show in which they were engaged with the performers; and producing a coherent and entertaining broadcast output) were met by *OOTW*. This assessment 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 a companion paper (Bowers, 1999), 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.

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

Our overall sense is that we succeeded in staging a gameshow 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 play able 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]
>
> I couldn't understand the frogs. I couldn't see what my team were doing.[inhabitant]

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—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 and the design principles described above. 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.

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 to support the cooperative work of TV production. 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. 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.

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.

> I had problems identifying with an avatar. It's the expressions and gestures which are missing.[audience]
>
> 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]
>
> I was straining to hear what people were saying so I didn't want to make a lot of noise. [audience]
>
> 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 amore sparse, albeit coherent landscape that contributed to the feeling of emptiness.

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. However, it 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.

Format

Our choice of a gameshow 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.
>
> Why do a gameshow at all? It's something with a narrow age-range appeal. You should do something more imaginative.

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 gameshow:

> A paradox for a technology that promises openness.

Although we would justify the choice of the gameshow 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 gameshow 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.

Summary and future work

Inhabited TV aims to create a new entertainment and communications medium by combining traditional TV with CVEs so that the public can become on-line participants within TV shows. Our paper 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; and control of virtual cameras.

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 community and broadcasting.

The lessons from *OOTW* may be relevant to other areas of CSCW. *First*, 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 could 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 might occur 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 (a point sometimes not fully appreciated in the debates about 'formalisms' in CSCW).

Second, the idea of deliberately capturing and making 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. *Third*, our notion of object-centred navigation (here exemplified in the camera control interface) may offer a novel and generally applicable alternative to conventional 6DOF navigation in virtual worlds.

We are currently planning our next experiments in inhabited TV. 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 and directing facilities. We hope that it will then be possible to create innovative and engaging inhabited TV.

Acknowledgements

We acknowledge the support of the European Community for the *eRENA* project under the ESPRIT IV *Intelligent Information Interfaces* (I3) programme and also the EPSRC for the *Multimedia Networking for Inhabited Television* project under the *Multimedia Networking Applications* programme. Additional support *for Out of this World* was provided by FACT and North West Arts.

The Mirror was a collaborative project with Sony and the BBC, whilst *Heaven and Hell – Live!* involved Sony and Channel 4. We gratefully acknowledge the involvement of many colleagues in those organisations, and also the contributions of numerous others at BT Labs, Nottingham University, Illuminations and the Royal Institute of Technology.

References

Barrus, J. and Anderson, D. (1996): 'Locales and Beacons.' Proc. *1996 IEEE Virtual Reality Annual International Symposium (VRAIS'96)*, San Jose, US, Kluwer.

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

Benford, S. D. and Greenhalgh, C. M. (1997), Introducing Third party Objects into the Spatial Model of Interaction, Proc. *ECSCW'97*,189-204, Lancaster, UK, Kluwer.

Bowers, J. (1999), Crossing the line: A field Study of an experiment in inhabited television. Submitted toECSCW99.

Capin, T. K., Pandzic, I. S., Noser, H., Magnenat-Thalmann, N., Thalmann, D. (1997), 'Virtual Human Representation and Communication in VLNet Networked Virtual Environments', *IEEE Computer Graphics and Applications*, IEEE Press.

CINEMATRIX (1998),http://www.cinematrix.com/

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

Greenhalgh, C. and Benford, S. (1995), 'MASSIVE: A Virtual Reality System for Tele-conferencing', *ACM Transactions on Computer Human Interaction*, 2 (3), pp. 239-261.

Lea, R., Honda, Y. And Matsuda, K. (1997), 'Virtual Society: Collaboration in 3D Spaces on the Internet', *Computer Supported Co-operative Work: The Journal of Collaborative Computing*, 6, 227-250, Kluwer.

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

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

McGrath, A., Oldroyd, A. and Walker. G. (1998), 'The Mirror: reflections on inhabited TV', Proc. *CSCW'98 Video programme*, Seattle, ACM Press.