The i2tv System: A Mixed Reality Communication Interface

Jasminko Novak jasminko.novak@gmd.de
Monika Fleischmann fleischmann@gmd.de
Wolfgang Strauss strauss@gmd.de
Predrag Peranovic peranovic@gmd.de
Christoph Seibert seibert@gmd.de

MARS – Exploratory Media Lab, IMK – Institute for Media Communication
GMD – German National Research Center for Information Technology
Sankt Augustin, D-53754, Germany

Abstract

This paper is concerned with developing the concept of interactive Internet-TV (i2tv) as a basis for medial integration of Internet participants into events taking place at a real physical location. Interaction in real space is combined with actions of Internet participants into new forms of production, such as interactive Mixed Reality TV. The paper discusses two main issues: (1) the development of the i2tv system as a technological framework for concepts integrating on-line and on-site participation, (2) two concrete dramaturgical models and experiences with public trials. The innovative aspect of the i2tv approach is the integration of Internet-based multi-user interaction and awareness with broadcast technologies such as Internet streaming and Virtual Studio, and with technologies for mixed reality in shared physical space.

1. Electronic arenas

The promise of interactivity is that the experience of cultural production can be something the viewers do rather than something they are given. This requires rethinking conventional concepts of content. In interactive media the interface decisively shapes the experience of content by defining how one perceives and navigates that same content. A number of different approaches focusing on real-time participation of remote participants in „media-rich" cultural events can be referred to under the term "electronic arenas" [25].

Approaches demonstrating this understanding of electronic arenas include Inhabited TV, Mixed Reality Stage and TV shows integrating contributions of Internet participants on one hand, and a range of approaches to shared virtual environments (CVEs, MUD/MOOs, etc.) on the other. Inhabited Television is based on broadcasting interaction from collaborative virtual environments (CVE) on TV. The action of participants within the virtual world is transmitted to a conventional viewing audience, either as a live event or sometime later as edited highlights. CVE technologies (e.g. MASSIVE-2, MS Virtual Worlds) are combined with broadcast technologies [3] [9].

The Mixed Reality Stage [18] is a model for electronic arenas bringing together several participants in physical space with participants from the Internet, in a combination of shared physical and three-dimensional virtual space. The virtual space is realised as an interactive field of audio-visual objects, which are triggered by users movement and emitted into the physical space. The Mixed Reality Stage is based on the e-MUSE system (Electronic Multi-User Stage Environment) [18] [7].

Many television transmissions involve the audience through questions and comments by telephone, fax or Internet (e.g. CNN's *Q&A* [26]). Kulturserver [14] combines Internet streaming and TV broadcast by feeding users' streams from the Internet (live performances or pre-produced material) into the Berlin Open Acces cable channel, accompanied by a simultaneous chat of the Internet viewers.

2. The i2tv system

i2tv (interactive Internet-TV) is a basis for medial integration of Internet participants into events taking place at a real physical location: a symposium, conference or live artistic production. The challenge is to create a situation in which on-line and on-site participants feel present and involved, while retaining the specifics of both situations (on-line, on-site).

To achieve this the i2tv system combines technologies for multi-user interaction and awareness with broadcast technologies such as Internet streaming and digital TV, and with technologies for

mixed reality in shared physical space. It is built as a modular system providing independent levels of implementation for broadcasting the live situation on-site to on-line participants, for supporting presence and interaction of on-line and on-site participants, for different input and display devices, and for medial staging on-site. It consists out of the following parts:

- MOO server – platform for networked multi-user environments [5][13],
- RealServer – platform for streaming video over Internet [23],
- e-MUSE – system for multi-user interaction in a combination of shared physical and virtual space [18],
- 3DK Virtual Studio – distributed Virtual Studio system [20][1],
- eMOOSE – interface layer connecting individual elements of the i2tv system.
- Display and input devices – Web browser, handhelds, free body interfaces.

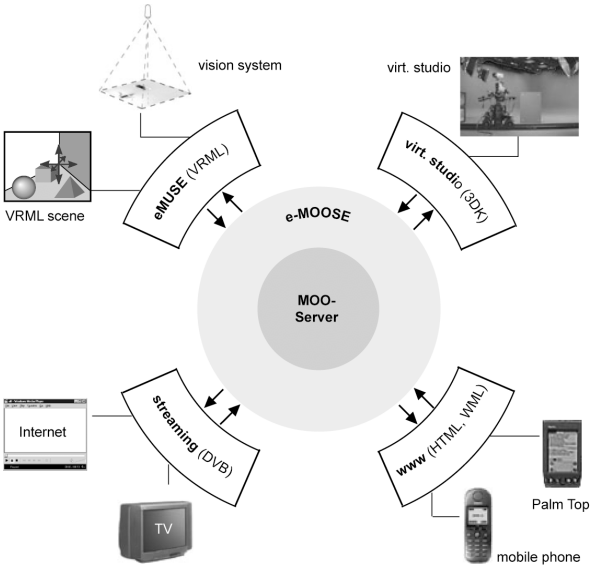

Figure 1. Basic structure of the i2tv system

The MOO system [5][6][10] provides support for multi-user communication and awareness, event propagation, spatial structuring, behaviour scripting and shared virtual space. The i2tv system extends the text-based MOO system into a framework for spatial structuring of hypermedia elements. The MOO is employed as an abstraction layer for generic representation of a shared virtual space, independently of the format of content representation (text, 2D, 3D). It serves as a universal layer for modelling interactions between on-site and on-line participants.

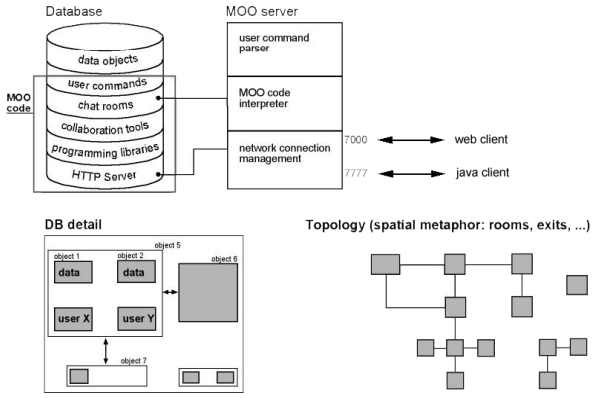

Figure 2. Structure of the MOO system in i2tv

This is fundamentally different to 3D virtual environments, which are built on structures describing the rendering of the shared space as the primary element. In contrast, i2tv builds on a layer describing data structures, interaction models and behaviours, independently from the layer describing how the content is to be rendered.

The integration of the e-MUSE system [18] in i2tv enables the deployment of VRML-based virtual spaces combined with multi-user interaction in a shared physical space, and the attachment of various interfaces such as vision systems, electric field sensors and handhelds.

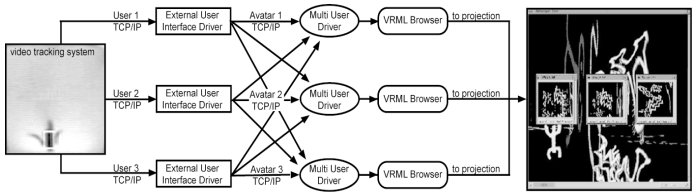

Figure 3. Structure of e-MUSE system

The 3DK distributed Virtual Studio system[1] [20][4] enables the composition of video images of a real broadcast camera together with 3D computer generated scenes, as well as the insertion of virtual objects into real scenes, with individual facilities distributed over an ATM network. It is a part of the Digital Video Production (DVP) system (Fig. 4), which consists of a wide area ATM testbed connected to digital video production test sites with support for broadband distribution of digital audio/video applications, and with distributed video production applications (e.g. distributed virtual

[1] Developed by the DMP research group at GMD's Institute for Media Communication.

studio, distributed video editing or simulation system).

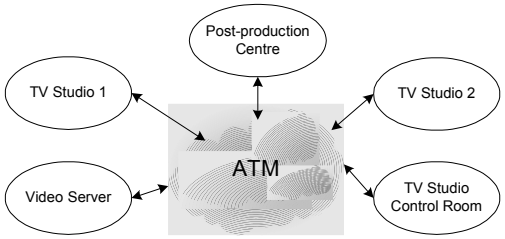

Figure 4. Digital Video Production System

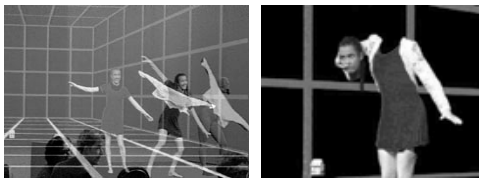

Figure 5. Compositing real actors with a 3D virtual set using Virtual Studio

e-MOOSE is a java-based interface layer connecting the individual elements of the i2tv system (MOO server, streaming, e-MUSE, 3DK, displays and input devices) into an integrated whole. It consists of a server-side java application and client-side applets. The client-side applets communicate with display systems and with interfaces to users‘ input devices. The server-side component communicates with the MOO server and with the client-side applets of individual users. The eMOOSE layer also takes care of mapping the MOO events into events suitable for a different display formats and input devices, and vice versa.

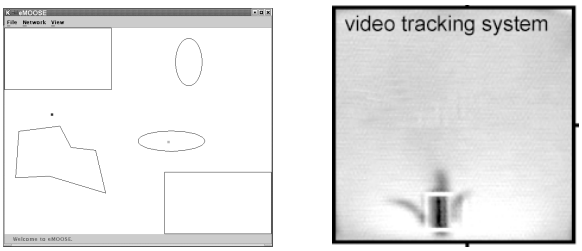

Figure 6. eMOOSE GUI for Spatial Region Mapping

To achieve this e-MOOSE interprets user's actions with her input device of choice as MOO commands, and maps the MOO virtual space and events to the user’s display system of choice. A working example of this translation is using body movement in real space as a means of interaction and navigation in a virtual space modelled as MOO topology. A vision system tracks the movement of an object (the user's body) in real space and transmits the position information to e-MOOSE. Using the eMOOSE Spatial Region Mapping Interface, the real space can be divided into regions that are assigned to selected areas of the MOO topology. Entering and leaving these areas triggers events that can be assigned to arbitray commands and behaviour scripts in the MOO. In this way different forms of connection between real and virtual space can be modelled.

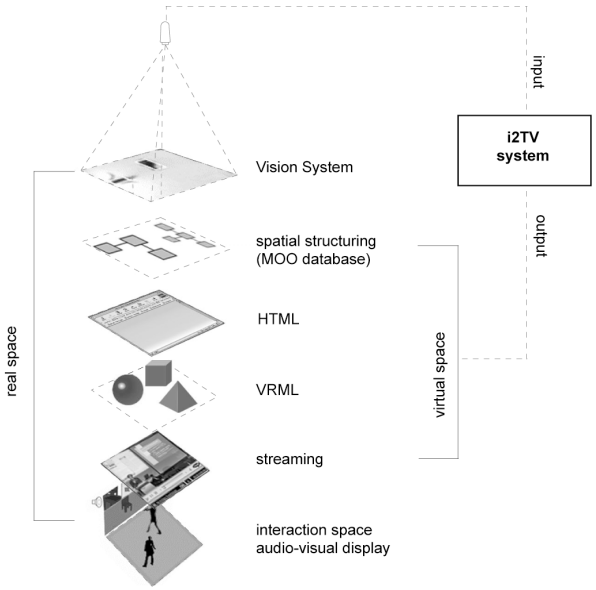

Figure 7. Linking different forms of real and virtual space in the i2tv system

The described multi-layered architecture of i2tv enables the combination of different levels of content representation, interaction and communication channels, based on the situations on-site and on-line, and depending on bandwidth, input and display devices of individual users. This distinguishes i2tv from other systems such as 3D shared virtual environments (e.g. MASSIVE) tele-conferencing systems (e.g. SunForum, MS Flatland), or web-based distance education systems (e.g. MS Telep).

2.1. Broadcasting the situation on-site to participants on-line

The first task in on-site/on-line scenarios is to provide feedback of the situation on-site for on-line participants. A straightforward way to do this is video streaming. While technologies such as Real Streaming enable reaching relatively large audiences with relatively low bandwidth requirements, they inevitably introduce 15-20 seconds delay between the live situation on-site and the video stream received by

on-line participants. This imposes limitations on the models of interaction between on-line participants and the situation on-site.

A possible solution is to use Internet as interaction channel, while broadcasting the on-site situation through TV. The i2tv system supports both real streaming for Internet, as well as satellite DVB (MPEG-2) streams, which has been demonstrated in the public trial at the *Memoria Futura* Symposium (Section 3.1). Upcoming digital TV set-top boxes often integrate a modem for Internet-based interaction back channel. This could provide basic (and limited) means of interaction for participants using digital TV both as display and interaction device. In this respect, especially interesting is the upcoming standard for the Multimedia Home Platform.

The current i2tv streaming solution is based on the Real Server (for low-end) and Kasena MediaBase (for MPEG-1, MPEG-2 quality). Since the streaming player runs as part of the i2tv web-interface, it is connected to the underlying MOO networked environment through the eMOOSE interface layer. This enables the implementation of scenarios based on events caused by individual user‘s actions upon the video. All these events can be propagated within the i2tv system, and used as triggers for events in other users clients. This enables the realisation of scenarios such as dynamic shared annotation spaces, shared video awareness, user video activity visualisation etc. It also enables the attachment of meta content to the video stream, stored in a database and synchronised to streamed content based on pre-defined time-markers or external actions (moderator, users).

2.2. Channels of presence & interaction

The i2tv system supports scalable channels for representation of remote users. In order to enable appropriate forms of interaction for different roles of on-line and on-site participants, on-line viewers, and on-site audience, a range of interaction channels is supported:

- moderated chat (public and private channels),
- streaming audio/video,
- text annotations to live video (video notes),
- 3D content creation in Virtual Studio,
- movement and content creation in shared 3D space.

Because of its low disruption effect text chat has proven the most viable form of action for on-line participants in the context of an integrated on-line/on-site discussion format such as a conference or symposium. It is also the only currently possible direct form of interaction for participants using handheld devices such as palmtops or cellular phones. This form of interaction is directly enabled by the MOO system.

Depending on available bandwidth and personal choice [12], the users can also choose between different forms of representation: live video stream, iconic representation through a photo or a symbolic graphical representation. The video stream from a user’s webcam is distributed to other users through the real server in the i2tv system. Since i2tv integrates the display of streaming video with the underlying multi-user event model, it is possible to combine events based on actions of individual users with manifestations in the display of incoming streams. For example, quieting down one of the users streams by the moderator can be automatically propagated to all on-line participants, affecting their display of the specified user stream.

In order to provide an appropriate channel of active participation for on-site audience and remote viewers i2tv introduces the possibility of live video stream annotations. The audience on-site and viewers on-line can make annotations to the video stream of the live situation, using their cell phone, palmtop or home PC. They assume the role of commentators. The comments are stored in the i2tv system as annotations to the video stream being recorded. Instead of a linear timeline, the video recording is visualised according to the timepoints set by the viewers’ actions. They provide the structure onto which all the individual elements of the event are mapped: the recorded video, the inputs of on-line participants, and the comments of on-line viewers and on-site audience.

Putting these individual pieces into relation provides a kind of a mental map of the event, as it is perceived by the participants involved. Instead of an archive of “objective” facts, the event is visualised as a collection of personal impressions.

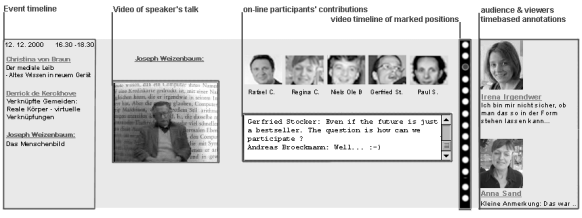

Figure 8. Visualising the event as a collection of personal impressions

This allows the participants to experience points of view different from one's own, while the event is still in progress. The relevance of such interaction channel for audience participation is supported by a recent study of video annotations in collaborative video viewing sessions [2] published parallel to our work. To the best of our knowledge, there have yet been no demonstrations and studies on using video annotations as a participation channel during a live event, as we intend with the i2tv system.

Perhaps the most interesting aspect of the i2tv model is the possibility to combine different interaction channels with each other. Even the simplest text inputs gain a different perspective if visualised as elements of the Virtual Studio setting, as is demonstrated in the „Ottos Mops" distributed play (section 3.2). Introducing free movement in physical space as the natural form of interaction for on-site participants creates a new layer of the performative situation. Connecting bodily action in physical space and on-line participation in shared 3D space enables different forms of Mixed Reality performances [18][7].

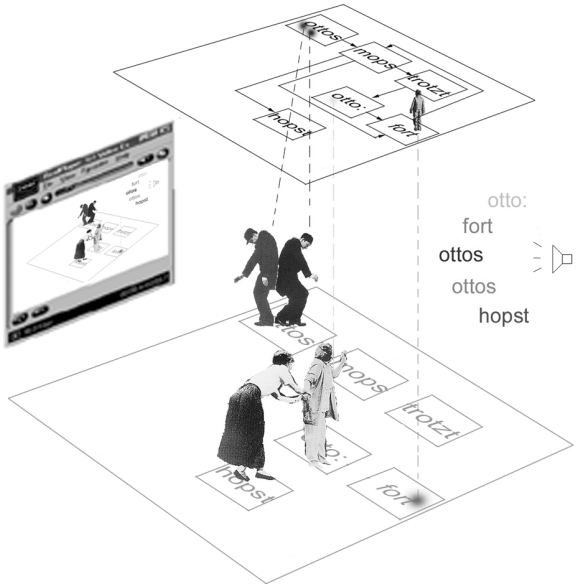

Figure 9. Combining different channels of interaction and representation

2.3 Organising many simultaneous inputs: medial staging

The basic characteristic of integrating on-line participants into an on-site event is the resulting multitude of simultaneous inputs that need to be organised into some kind of integrated whole. One approach to this problem is clustering individual inputs into larger entities, where manifestation is varied based on the number of participants involved in the same kind of action (e.g. crowd management techniques [3]). In our approach we look at theatrical staging as an organisational principle for many simultaneous information sources. This requires appropriate technologies for spatialised audio-visual displays. The common approach is to have spatially organised projections of shared 2D or 3D spaces. We have explored this approach in our previous work on developing the Mixed Reality Stage [18].

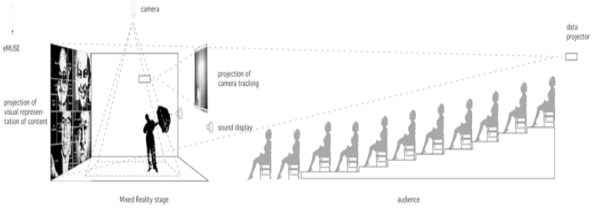

Figure 10. Example of spatial setup on-site using projection VR [18]

The Mixed Reality Stage employs an interactive sound space for spatial immersion, accompanied by visual projections as secondary elements. This is one workaround to the current impossibility of visually constructing a virtual 3D space within physical space, due to the lack of real-time holographic displays.

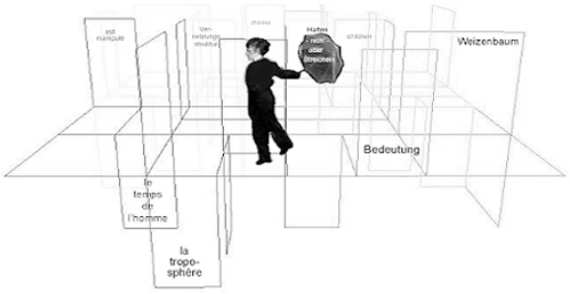

Figure 11. The principle of interactive soundscape

Another possibility is using glasses with see-through personal displays to which stereoscopic images are transmitted from a remote server. However, both of these approaches are difficult to convey to on-line participants and not suitable for broadcast format. We resolve this problem by using the Virtual Studio technology as a display environment for Mixed Reality scenarios. This is demonstrated by the „Ottos Mops" distributed play where the virtual stage is used to spatially organise the on-line inputs (Fig. 16). A final possibility is to use CAVE-like systems such as GMD's Cyberstage. This would provide the advantage of immersive experience for an on-site performer but due to the limited-size of the CAVE (3mx3mx3m) only a limited stereoscopic experience for a small on-site audience. A related experiment in creating a CAVE-based networked performance environment is described in [24]

3. Demonstrators and public trials

As part of developing the i2tv system two demonstrators have been produced and employed in public trials.

3.1. Memoria Futura Symposium

The i2tv trial at the Memoria Futura Symposium[2] in December '99 at GMD worked with the common conference format as a point of departure in developing new models for connected on-line/on-site events. The goal of the experiment was to explore a minimal set of requirements needed to integrate on-line participants into the situation on-site. To achieve this, a group of invited experts was integrated into the discussion on-site as Internet participants through the minimal version of the i2TV system. Live audio and video from the symposium were streamed to remote participants as Internet streaming and as digital TV broadcast via satellite.

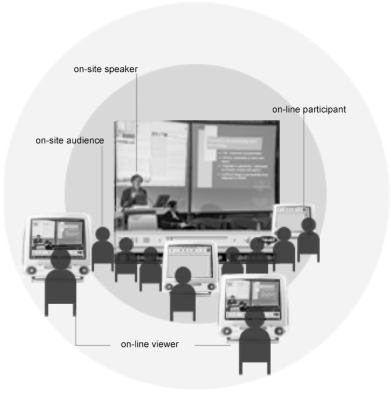

Figure 12. Layers of participation in the i2tv trial at *Memoria Futura* Symposium

On-line participants could intervene into the symposium by means of text, which was made visible on-site by projecting the Internet view into the real space of the symposium. Additionally, most important questions from the online participants were selected by the on-line moderator and displayed on the fly in large letters. These selections were also propagated to the web interface of individual on-line participants in order to provide a feedback of the moderator's actions.

[2] "Memoria Futura: Information Technology and Cultural Heritage – A New Perspective?" , Dec. 11-12, 1999 , GMD, Schloss Birlinghoven, Germany. Four thematic sessions combined four on-site speakers and five invited on-line participants per session. On-line participants connected from Germany, Italy, Denmark, Croatia, and the USA. http://imk.gmd.de/mars/cat/memoria

The technical realisation of this restricted model combined three basic elements of the i2tv system: the MOO server, the Real Server, the prototype eMOOSE client and the web interface.

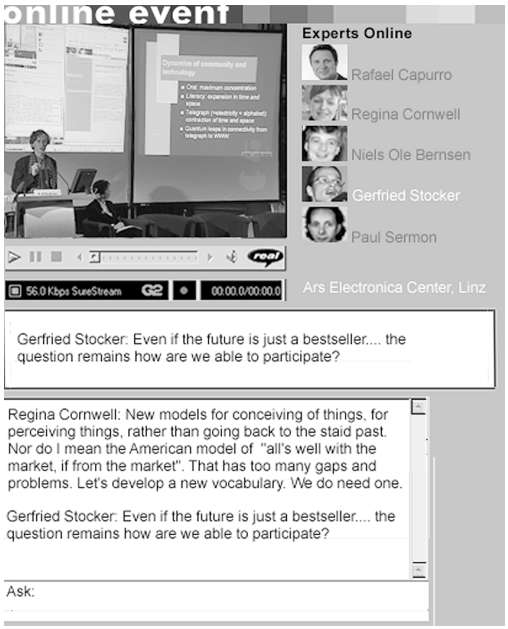

Figure 13. Minimal i2tv interface at Memoria Futura Symposium

The most important lesson of the Memoria Futura trial[3] is understanding the four specific roles inherent to connected on-line/on-site situation:

- on-site speaker: the actor,
- on-site audience: the passivly involved,
- on-line participants: the commentators,
- on-line viewers: the analysts.

The complaints of on-site audience as being disadvantaged to on-line participants shows that available channels for active participation can be more important than physical presence.

3.2. Distributed play „Ottos Mops"

„Ottos Mops" is a model for a live artistic production integrating on-line participants with participants at a real physical location. It exemplifies an on-line/on-site scenario in which on-line participants are an integral part of the situation on-site and crucially determine the content produced. "Ottos Mops" brings together an on-site performer and multiple on-line participants in a real-time distributed play based on sound poetry of the Austrian poet Ernst Jandl.

[3] For detailed analysis and evaluation of the trial see [17].

Jandl's phonetic poem is split into individual words in this Mixed Reality television game. Participants on-site, in Internet cafés or outside in the street, are equipped with mobile phones, palmtops, PCs or touch screens, in order to send their word contributions into the Virtual Studio. The moderator in the studio becomes a conductor who binds the incoming words into a new phonetic poem. Beyond the real-time interaction, the individual contributions of the participants are related to each other and visualised in the form of a Memory Space.

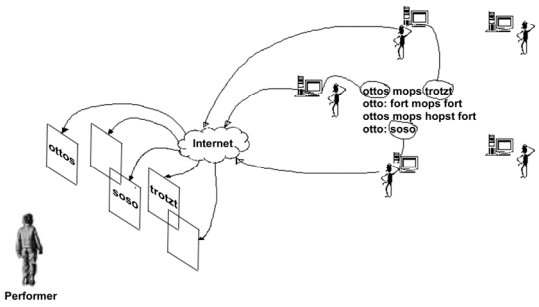

Figure 14. Basic Concept of "Ottos Mops"

The realisation of *Ottos Mops* deploys the following i2tv configuration: the MOO server, real streaming, eMOOSE, WWW, and Virtual Studio. On-line participants and viewers can use the web interface both for interaction and display, or watch the TV broadcast for high-quality video while using the web interface only for interaction.

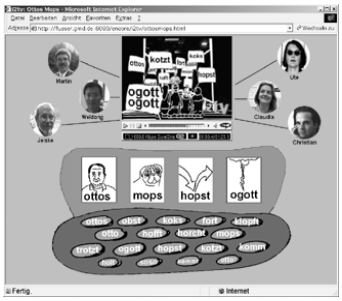

Figure 15. i2tv web interface for "Ottos Mops"

The on-site performer is integrated into the virtual stage by using real-time Virtual Studio technology. The output of the Virtual Studio is either streamed to Internet audience or broadcasted to TV audience. In this way, the on-line participants can experience the situation of a Mixed Reality Stage created in real-time through their interaction with the participants on-site in the studio. The audience on-site can both see the actor in the blue box on the real stage and watch the projection of the resulting Mixed Reality stage, with the combined audio inputs from the actor and from on-line participants.

Figure 16. Medial staging of on-line inputs in Virtual Studio for "Ottos Mops" distributed play

Staging *Ottos Mops* as a distributed play can be understood as creating a situation in which the participants develop strategies for exploring interactive narratives. This concept points to possible approaches for developing participation models for networked scenarios which go beyond traditional models such as game shows, quizzes, or social chats in 3D space. Rather than merely a combination of underlying technologies, *Ottos Mops* demonstrates the abstraction of different cultural models from everyday life (TV, Internet) into an artistic production.

4. Conclusions

In this work, we have developed, applied and evaluated the i2tv system as an electronic arena for participation in cultural events integrating Internet and physical space. The i2tv system combines networked multi-user interaction and awareness with medial staging on-site and with broadcast technologies such as Internet streaming and Virtual Studio.

Rather then placing the participants as avatars in a shared virtual world, or creating a mixed reality performance confined to the physical space of the theatre, the i2tv system enables the development of networked Mixed Reality productions linking physically present and remote participants while retaining the specifics of their respective situations. The two i2tv demonstrators explore new forms of cultural production and participation. The historical format of public discussion is extended to integrate on-site and on-line participants into a networked Mixed Reality situation at the Memoria Futura Symposium. This exemplifies a point of departure for uncovering the intrinsic characteristics of the connected on-line/on-site situation as a fundamentaly new form of experience.

The distributed play *Ottos Mops* represents a theatrical model integrating on-line participants as

active producers of new content and an integral part of the situation on-site. It demonstrates a participation model based on four roles specific to the connected on-line/on-site situation: the actor (on-site performer), the passively involved (on-site audience), the commentators (on-line participants) and the analysts (on-line viewers). Realising a live Virtual Studio production allows to visualise a Mixed Reality situation for all parties involved, while reaching large audiences through live TV broadcast. The combination of the passive broadcast model with interactive channels of the i2tv system enables active participation of the audience through the Internet interface. This combination demonstrates the integration of the participatory model of the Internet and the passive broadcast model of TV into a new convergent unit - live Mixed Reality TV production.

References

[1] http://imk.gmd.de/dmp

[2] Bargeron, D.; Gupta, A.; Grudin, J.; Sanocki, E; Annotations for Streaming Video on the Web: System Design and Usage Studies, Proc. WWW8, 61-75

[3] Benford, S. et al.; Evaluating Out Of This World: An Experiment in Inhabited Television, In Second Year eRENA Deliverables, 1999, Stockholm, ISBN 91-7170-466-3

[4] Breitender, C. et al.: Distributed Video Production on ATM, Proceedings of ECMAST '96, April 1996

[5] Curtis, Pavel and David A. Nichols, "MUDs Grow Up: Social Virtual Reality in the Real World," Proceedings of the 1994 IEEE Computer Conference, pp. 193--200, January 1994

[6] http://lingua.utdallas.edu/encore/

[7] Fleischmann, M.; Strauss, W.; Novak, J.; Murmuring Fields Rehearsals -Building up the Mixed Reality stage, Proc. of KES 2000, Brighton, Aug. 31-Sept. 2

[8] Benford, S. D.; Greenhalgh, C. M.; A Multicast Architecture for Large-Scale Collaborative Virtual Environments, in Proc. of ECMAST'97, Milano, Italy, May 1997, pp.113-128

[9] Greenhalgh, C.; Benford, S.; Taylor, I.,; Bowers J.; Walker G.; Wyver J.; Creating a Live Broadcast from a Virtual Environment, in Proc. ACM SIGGRAPH'99 pp.375-384, ACM Press

[10] Haynes, C.; Holmevik, J.R. (eds.); High Wired: On the Design, Use, and Theory of Educational MOOs. Ann Arbor, Michigan, University of Michigan Press, 1998

[11] Isaacs, E.A.; Morris, T.; Rodriguez, T.K.; A Forum for supporting interactive presentations to distributed audiences, Proc. CSCW '94, 405-416, ACM

[12] Jancke, G.; Grudin, J.; Gupta, A.; Presenting to Local and Remote Audiences: Design and Use of the TELEP System, Proc. CHI 2000, 384-391

[13] ftp://ftp.lambda.moo.mud.org/pub/MOO/

[14] http://www.kulturserver.de/

[15] Novak, J.; Development of Paradigms and Models for Interactive Environments, Master Thesis, Faculty of Electrical Engineering and Computing, University in Zagreb, July 1999, and GMD Research Series, ISBN 3-88457-372-1

[16] Rekimoto, J.; Ayatsuka, Y.; Uoi, H.; Arai, T.; Adding another communication channel to reality: An experience with a chat-augmented conference. CHI'98 Summary, 271-272.

[17] Novak, J et al.; An electronic arena for participation in connected communities, GMD eRENA 3rd Year Deliverables, GMD, August 2000, ISBN 91-7170-611-9

[18] Strauss, W.; Fleischmann, M. et al.; Linking between real and virtual spaces, In eRENA 2nd Year Deliverables, 1999, Stockholm, ISBN 91-7170-466-3

[19] Vellon, M. et al.; The Architecture of a Distributed Virtual Worlds System, Microsoft Research, 2000

[20] Vonolfen, W.; Virtual Studios for TV Productions – Limits and Chances, Proceedings of the eight IEEE international workshops on enabling technologies: Infrastructure for collaborative enterprises, Stanford, USA, June 16-18, 1999

[21] Strauss, W.; Fleischmann, M. et al.; Staging the space of mixed reality - reconsidering the concept of a multi-user environment, Proceedings of VRML 99, 1999, Paderborn, Germany, pp. 93-98

[22] White, S.; Gupta, A.; Grudin, J.; Chesley, H.; Kimberly, G.; Sanocki, E.; Evolving Use of a System for Education at a Distance, Proc. HICSS-33, CD-ROM

[23] http://www.realnetworks.com/

[24] Sengers. P.; The avatar as interface in virtual environments, GMD eRENA 3rd Year Deliverables, GMD, August 2000, ISBN 91-7170-611-9

[25] http://www.nada.kth.se/erena

[26] http://www.cnn.com/q&a