

Report on Final Demonstration



FascinatE identifier: Fascinate-D631-BBC-Report on Final Demonstration-v04.docx

Deliverable number: D6.3.1

Author(s) and company: G. Thomas (BBC); O. Schreer (HHI);
G. Thallinger, G. Kienast (JRS); R. Oldfield (UOS);
J. Ruiz-Hidalgo (UPC); J.-F. Macq (ALU);
M.J. Prins (TNO)

Internal reviewers: G. Kienast (JRS), L. Barkhuus (TII)

Work package / task: WP6 / T6.3

Document status: Final

Confidentiality: Public

Version	Date	Reason of change
1	2013-06-10	Document created (structure proposed, initial input)
2	2013-06-16	Input integrated from HHI, UPC, UOS, TNO; other additions made
3	2013-06-19	Additional input from ALU and other additions following discussions at consortium meeting. Version for internal review
4	2013-06-21	Review comments addressed. Final version.

Acknowledgement: The research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 248138.

Disclaimer: This document does not represent the opinion of the European Community, and the European Community is not responsible for any use that might be made of its content.

This document contains material, which is the copyright of certain FascinatE consortium parties, and may not be reproduced or copied without permission. All FascinatE consortium parties have agreed to full publication of this document. The commercial use of any information contained in this document may require a license from the proprietor of that information.

Neither the FascinatE consortium as a whole, nor a certain party of the FascinatE consortium warrant that the information contained in this document is capable of use, nor that use of the information is free from risk, and does not accept any liability for loss or damage suffered by any person using this information.

Table of Contents

1	Executive Summary	1
2	Introduction	2
2.1	Purpose of this Document	2
2.2	Scope of this Document	2
2.3	Status of this Document	2
3	Demonstration of whole System during a live Performance	3
3.1	Overview of Live Demonstration System	3
3.2	Immersive Experience over a Full-Bandwidth Link with User Control	6
3.3	Interactive UHDTV Meets Network	7
3.4	My Personalised Production	8
4	Stand-alone Demonstrations	9
4.1	OmniCam	9
4.2	Broadcast Camera Integration	11
4.3	Virtual Director	12
4.4	Content Analysis	13
4.5	Delivery using Adaption in the Network	14
4.6	Delivery using Tile-based HTTP Adaptive Streaming	15
4.7	Gesture Control	16
4.8	Video Rendering	17
4.9	3D Audio	17
4.9.1	Audio Object Extractor	18
4.9.2	Higher Order Ambisonics	19
4.9.3	Wavefield Synthesis	20
5	Coverage in the Press	21
5.1.1	Book of the Future	21
5.1.2	One Publication	22
5.1.3	Manchester Gazette	22
6	Reflections on the Demonstration	24
6.1.1	Visitor Feedback	24
6.1.2	Further Technical Details of the Demonstrations	25
6.1.3	Acknowledgements	25
7	References	26
8	Glossary	27
9	Appendix 1: Flyer for Event	28
10	Appendix 2: Event Guide	30

1 Executive Summary

The objective of WP6 in the FascinatE project is to organise a series of convincing demonstrations that raise awareness of the project in the broadcast and media industry, as well as providing focal points for the technical work of the project.

This document reports on the third and final public demonstration of FascinatE technology, held at MediaCity UK, at the premises of the University of Salford. The centrepiece of the demonstration was the use of the end-to-end FascinatE chain being used to capture, deliver and display a live music and dance performance staged in the University's Digital Performance Lab. The performance ran three times during the day, with each show being preceded by a 30-minute presentation to introduce the project and explain the various aspects of the technology that were about to be demonstrated. This was accompanied by a set of stand-alone demonstrations that ran throughout the day, giving more in-depth insights into various results from the project. The event also resulted in several press publications, which are also listed in this deliverable.

During the demonstrations, audio and video data were captured to support the evaluation tasks during the remainder of the project, and for use for research beyond the end of the project.



2 Introduction

2.1 Purpose of this Document

This is the third of three documents reporting on the public demonstrations held by the FascinatE project.

The first public demonstration, held at IBC in September 2011 [D6.1.1] showed initial versions of some of the components that the project developed in its first year, but mainly as isolated elements, and not being used for a live production.

The second public demonstration of FascinatE technology was held in May 2012 [D6.2.1], at Arena concert hall in Berlin, during the production of a dance project by the Compagnie Sasha Waltz & Guests. Supporting demonstrations were shown at HHI's premises nearby. In contrast to the first public demonstration, the second demo showed some of the project's technology being used to capture a live event.

This document reports on the third and final public demonstration of FascinatE technology, held in May 2013 at MediaCityUK, at the premises of the University of Salford. The centrepiece of the demonstration showed the end-to-end FascinatE chain being used to capture, deliver and display a live music and dance performance staged in the University's Digital Performance Lab (described in Section 3). The performance ran three times during the day, with each show being preceded by a 30-minute presentation to introduce the project and explain the various aspects of the technology that were about to be demonstrated. This was complemented by a set of stand-alone demonstrations running throughout the day, which gave more in-depth insights into various results from the project (described in Section 4). The flyer produced to promote the demonstrations is included in Appendix 1, and the guide given to attendees is in Appendix 2.

2.2 Scope of this Document

This document is related to Task 6.3: Final demonstration.

2.3 Status of this Document

This is the final version of D6.3.1.

3 Demonstration of whole System during a live Performance

The live demonstration was based around the premiere performance of 'Deeper than all Roses', the latest large-scale music composition from Stephen Davismoon, featuring the band Bears?Bears!, in celebration of the works of the celebrated American poet E.E. Cummings.

Visitors were invited to view the performance being relayed live to a presentation room. The room capacity was limited to 20 people, so a total of 60 visitors were able to pre-book spaces. The demo was over-subscribed, so some visitors were only able to visit the non-live demos described in Section 4. The presentation room was equipped with a large projector and a spatial audio rendering system consisting of 17 speakers.

The live demonstration was divided into three phases, allowing specific aspects of the technology to be highlighted. The following sections describe the key technical elements of the demonstration system, and describe the three phases of the live demo.

3.1 Overview of Live Demonstration System

Behind the scenes, there were many items of equipment connected together to implement the various combinations of capture, coding, delivery, audio and video rendering and automated scripting that were demonstrated. The figure below shows all the elements of the system as set up for the demo. The main items of equipment and their interconnections are shown in Figure 1 below.

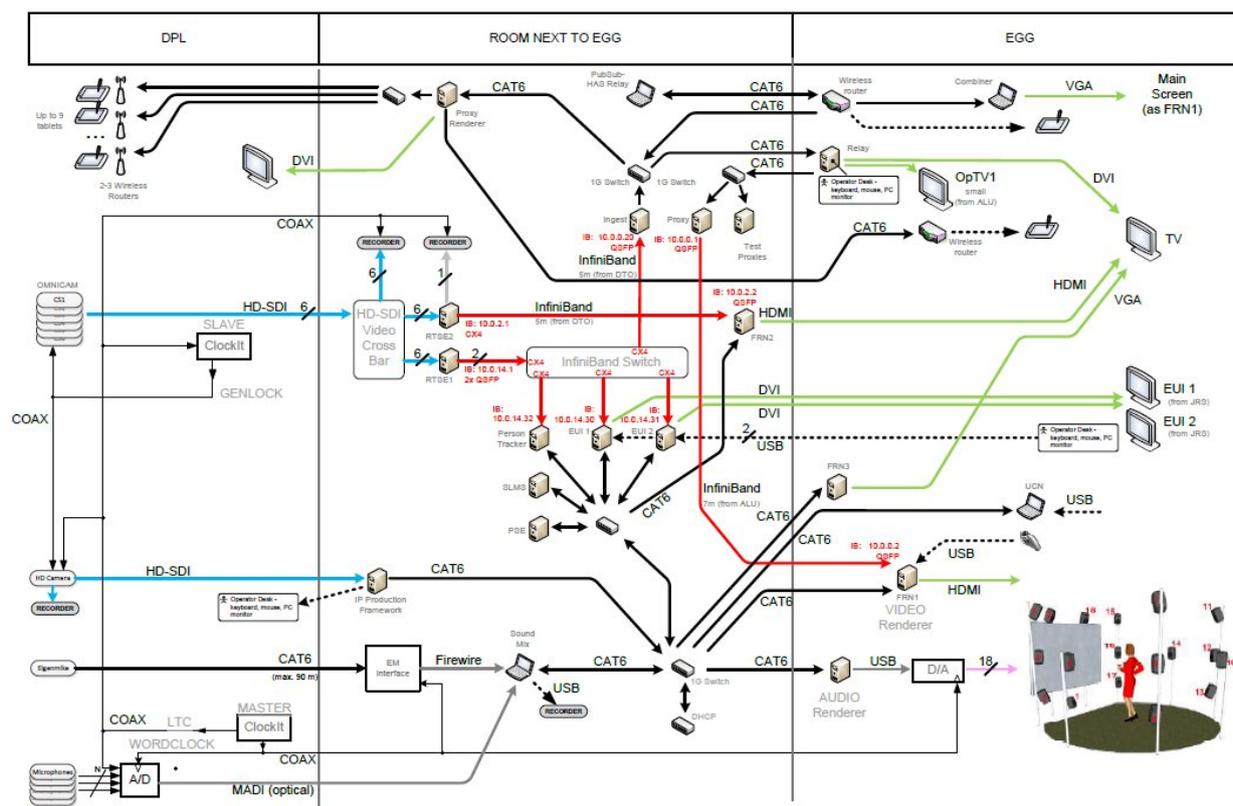


Figure 1: Wiring diagram for the live demonstration

The performance took place in the Digital Performance Lab (DPL), shown in Figure 2. The performance was captured with the OmniCam panoramic camera, using the latest version comprising of 10 HD cameras capable of giving a 360-degree panoramic image, of which 5 cameras were used for the demo to produce a 5k x 1k 180-degree live video panorama. A manned HD broadcast camera placed close to the OmniCam captured close-ups of particular areas of interest. The video from this was ingested into

an IP-based production framework, running a plug-in that generated camera metadata describing pan, tilt and zoom derived by tracking background features in the image.



Figure 2: Capture and performance area in the Digital Production Lab (DPL)

Audio was captured using an Eigenmike to generate a 4th order Ambisonic representation of the ambient sound in the performance room. Six additional audio objects were also captured using mics on the guitar amplifiers, drum kit and the PA speakers used by the singers.

The DPL also housed WiFi access points, providing network access to a set of tablet computers that visitors could use to explore the live panoramic image (see Figure 3).



Figure 3: Audience member using a tablet to navigate in the live panoramic video

The bulk of the processing equipment was placed in a room adjacent to the “Egg Suite” (the demo room where the presentations were given), shown in Figure 4. This equipment included:

- two Real-Time Stitching Engines (RTSE) that assembled a single panoramic image from the HD-SDI video signals from the cameras in the OmniCam (two were needed in order to provide sufficient output bandwidth to drive the other elements of the demo), and an Infiniband switch to distribute this high-bandwidth data;
- a PC running the IP Production Framework and camera tracker for the manned broadcast camera;
- a PC running the Ingest functions for the delivery framework, receiving the stitched OmniCam video signal from one of the RTSE's, and outputting compressed and segmented video streams via IP Multicast to 3 other nodes down the chain, using a 1GbE carrier-grade switch. At the other end of the delivery chain, a ‘Proxy’ PC re-assembles a subset of the video segments and output the reconstructed video to the FRN node (located in the Egg);
- a PC running a ‘Proxy renderer’ to process the pan-tilt-zoom navigation requests (render, encode and stream) for the 9 tablet computers used by audience members in the DPL;
- several PCs running elements of the content analysis and virtual director system: Editor User Interface (EUI), Semantic layer Management System (SLMS), Production Scripting Engine (PSE), Person tracker;
- recorders for the audio and video to produce data for later use.

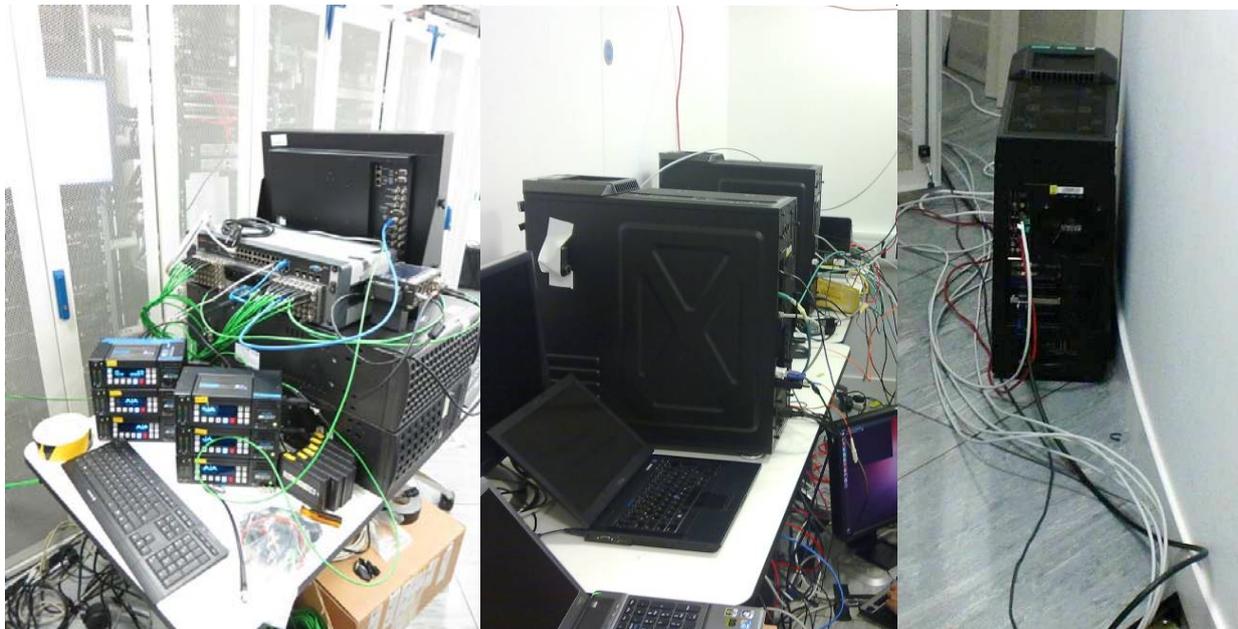


Figure 4: Equipment room: the Real-time Stitching engine, the Infiniband switch and the KiPro recorders (left); Proxy Renderer and Ingest with infiniband input (centre); Proxy with Infiniband output to FRN located in Egg Suite (right).

In the Egg Suite itself (Figure 5), there were PCs running several instances of the FascinatE Rendering Node (FRN) driving the main projector screen and an adjacent flat screen, a Relay node responsible for the adaptation of the tiled panoramic video streams, two terminals for the Editor User Interface (EUI) providing manual annotation and control of the Virtual Director system, and an audio renderer driving the 18-channel 3D audio system.



Figure 5: Demonstration area

Further details of the final integrated system will be documented in [D1.5.3]

3.2 Immersive Experience over a Full-Bandwidth Link with User Control

The first phase of the demonstration showed the output from the OmniCam (5k x 2k pixels at 25Hz) being streamed and rendered live by a video and audio renderer known as the FascinatE Rendering Node (FRN). The video from a conventional HD broadcast camera was also streamed live, and the corresponding area in the OmniCam image was highlighted, using data derived by real-time image-based tracking. The gesture interface was used to navigate around the panorama (Figure 6), for example allowing the display to focus on a guitarist of the drummer.

As the user panned around the panoramic image, the audio scene rotated accordingly, and when zooming, the audio objects were brought closer and the sound image widened to give the impression that the viewer was moving closer to the scene. The gesture-based control system was used to adjust the audio level and to mute/unmute it.

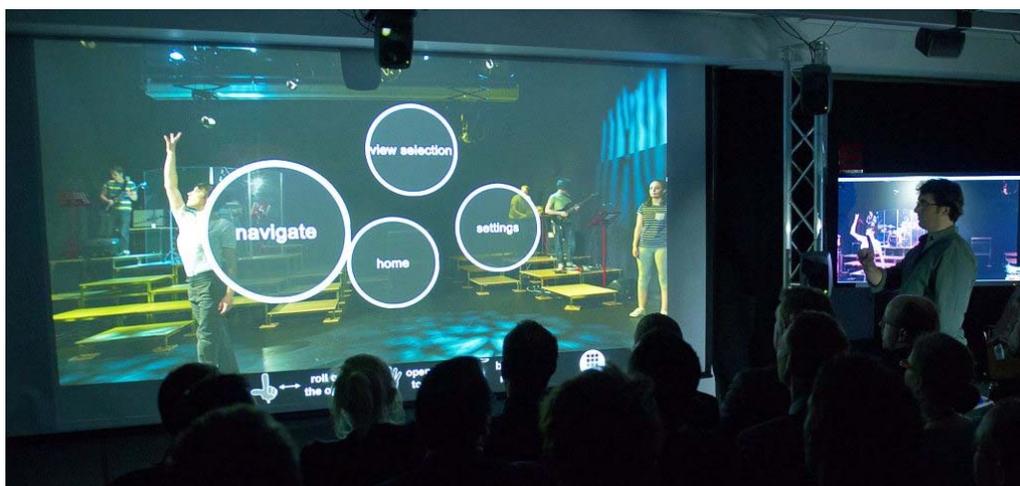


Figure 6: Demonstration of gesture-based interaction during the live performance

3.3 Interactive UHDTV Meets Network

The second phase of the demo showed how the immersive experience could be delivered over a lower-bandwidth network. The demonstration started by showing how poor the image quality would look if the whole panorama was downscaled to HDTV resolution, to emphasise the need for a more intelligent way of delivering an interactive panoramic image over a typical consumer network.

The demonstration went on to show how the FascinatE system tries to more intelligently react to the needs of the end-user, by selectively transmitting only the portions of the panorama that the user is looking at. This was achieved by representing the OmniCam content in an arrangement of independent video tiles (Figure 7). Downscaled versions were also produced to support different zooming levels. The image was divided into 85 tiles, which were encoded live in H.264 and pushed into the delivery system. A display showed which tiles were currently being delivered to the user, with the tile set adapting as the user navigated around. It was explained that this is analogous to browsing Google Maps, but now with video instead of still images, giving a video quality very close to the initial uncompressed case, while still streaming at a fraction of the total bitrate. It was also shown how tiles requested by several users could be sent using multicast rather than unicast. The demonstration was also shown on a tablet and smartphone in addition to the PC-based renderer, to show that a wide range of devices can be catered for.

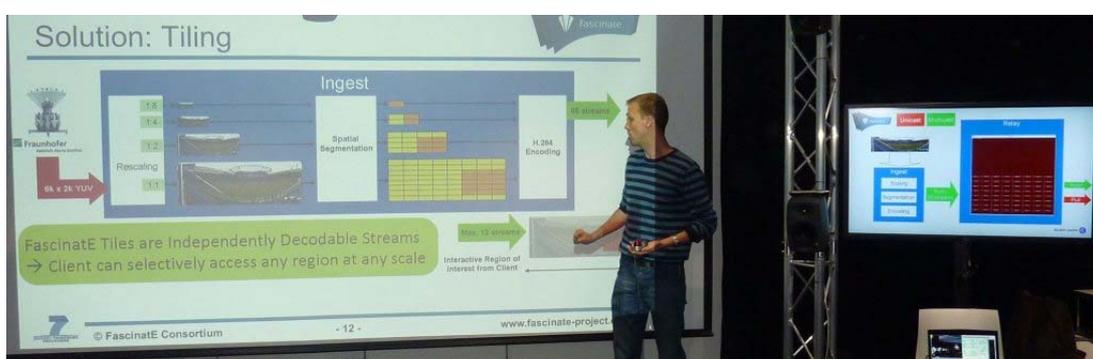


Figure 7: Demonstration of tile-based streaming

The final phase of the network part of the demo showed the project's extension to adaptive streaming (Figure 8), in order to support UHDTV content. This used the same multi-resolution tiled representation, as shown previously. In this case, though, the tiled video streams were each delivered using regular HTTP-based adaptive streaming, which means that standard Content Delivery Networks could be used for large-scale distribution. It was explained that this extension is now a candidate for standardization within MPEG. It was also shown how a tablet device could be used as a control for the PC-based renderer, providing another form of interactive control.

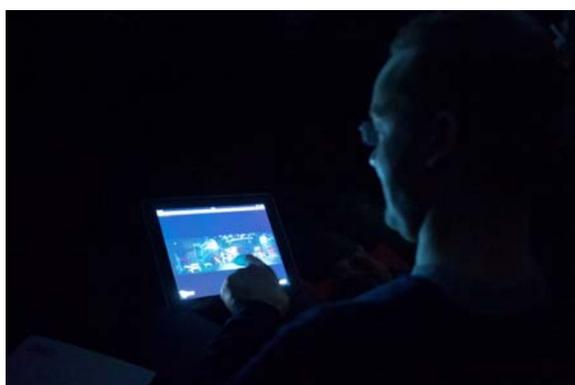


Figure 8: The performance delivered on a tablet via Tiled Adaptive Streaming

3.4 My Personalised Production

The final phase of the demonstration focused on the content analysis and automated director functionality that the project has developed. The aim of this is to allow multiple versions of a programme to be produced, with shot selection and framing being tailored automatically to each viewer's preferences and display size.

Several operators (sat on one side of the presentation room) used the content annotation tools to select visually-important areas of the panoramic scene (Figure 9). This information was used by two 'virtual director' systems to produce two outputs: one designed for viewing on the large display and another optimised for a smaller flat-panel display. It could be seen that both displays generally followed the main areas of interest (the dancers and the musicians) although the decisions of exactly how to frame the shots and when to switch between views varied slightly between the screens.



Figure 9: Content annotation tools being used live

4 Stand-alone Demonstrations

4.1 OmniCam

The OmniCam is a 180° panoramic video acquisition system. For the live demonstration, a new version of the omni-directional camera was used (see Figure 10). This new OmniCam has a much smaller form factor compared to the first version that was equipped with the Alexa M by ARRI. This new OmniCam is equipped with 10 Indie GS2K cameras mounted on a mirror rig with 10 mirrors capable of capturing a 360° cylindrical field of view. For the live demo, the video output from five cameras was used to create 180° panoramic video. The mirror rig with the cameras has a height of 60cm and a diameter of 50 cm. The weight is about 15 kilograms. The acquisition was performed at a frame rate of 25 frames/second.



Figure 10: The new OmniCam

The research leading to these results has received funding from the European Union's Seventh Framework Programme under grant agreement / 248138.

FascinatE – Panoramic Video Acquisition

The OmniCam – Two omni-directional cameras developed and used in project shootings

Ingo Feldmann¹, Arne Finn¹, Oliver Schreier¹, Christian Weisig¹, Wolfgang Walzenegger¹, Johannes Steurer²
[ingo.feldmann, arne.finn, oliver.schreier, christian.weisig, wolfgang.walzenegger]@hhi.fraunhofer.de, jsteurer@uarri.de
¹Fraunhofer Heinrich-Hertz Institute (HHI), ²Arnold & Richter Cine Technik

OmniCam by Fraunhofer HHI equipped with ARRI Alexa M

- High quality sensor, high dynamic range
- 6 cameras for 180° panoramic video
- Resolution of 6k x 2k
- Premium quality acquisition for immersive theatres
- Large sized and heavy weight (80 kg)

OmniCam by Fraunhofer HHI equipped with Indiecam

- Medium quality sensor
- 5 cameras for 180° panoramic video
- Resolution of 5k x 2k
- Medium quality acquisition for interactive services
- Small sized and light weight (15kg)

ARRI Alexa with separated head developed for the OmniCam

- Panoramic camera OmniCam requires a high quality sensor
- The ALEXA camera features high dynamic range, low noise, brilliant image quality, especially for difficult lighting (low light, high contrast, etc.), but still bulky and heavy
- Development of Alexa M with separated head to allow integration in the OmniCam
- Head with sensor separated to the body and connected by fibre

Real-time stitching engine for ultra-high definition panoramic video

- Capturing of 6 HD video streams up to 30 fps
- GPU and CPU based warping, stitching and blending of 6 camera views into one common 180° panoramic video at a resolution of 6k x 2k
- Processing is performed on a single PC + 3 DVS Atomix capture cards + 1 NVIDIA GTX 590 graphics card
- Simultaneous rendering of different regions at arbitrary size and resolution

The generation of panoramic video from multiple views

original views

photometrical correction - blending

geometrical correction - warping

final cropping

www.fascinate-project.eu

JOANNEUM RESEARCH
 Forschungsgesellschaft mbH
 DIGITAL – Institute for Information and
 Communication Technologies
 Heurystrasse 11
 8600 Leoben
 Austria
 Tel: +43 316 876 1240
 Fax: +43 316 876 1114
 info@joanneum.at
 www.joanneum.at/digital
 www.fascinate-project.eu

Figure 11: Poster for panoramic video acquisition demo

4.2 Broadcast Camera Integration

This demonstration showed how the video from a broadcast camera can be used in addition to a panoramic video feed to enhance the coverage of an event, using the example of a football match captured during an earlier test shoot. It explained how the broadcast camera can be calibrated with respect to the OmniCam, and how the depth in the scene of the main object of interest can be estimated in order to allow a switch between broadcast and OmniCam to occur without the viewer losing track of the object they are looking at, despite the parallax differences between the views. Other issues such as colourimetric alignment of the images were also explained.

The research leading to these results has received funding from the European Union's Seventh Framework Programme under grant agreement n° 258232.

fascinate

FascinatE –

Format Agnostic Script-Based Interactive Experience

Integration of broadcast camera and Omnicam

1. Motivation



Typically the Omnicam may be used in conjunction with one or more broadcast cameras:

- Large field of view
- Low angular resolution

Omnicam:

- Small field of view
- High angular resolution

The region of interest followed by the broadcast camera is defined by a camera operator. Thus we expect it to cover one of the most important parts of the scene, and to have very good shot framing and high resolution.

2. The Task

We want to integrate the video from the broadcast camera, and the region of interest it defines, into the FascinatE system. In addition to other regions-of-interest identified in the Omnicam image by the Virtual Director, we thus need to determine which region of the Omnicam image best corresponds to the broadcast camera view, present this information to an operator or end-user, and provide a means to transition between Omnicam and broadcast camera images that takes account of differences in the camera position.

1. Illustrating the region of interest viewed by the Broadcast Camera in the OMNICAM view.
2. Drawing the video from the Broadcast Camera at an appropriate place with respect to the OMNICAM.

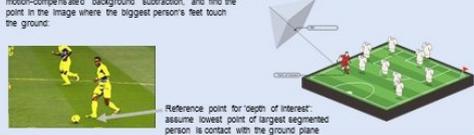


The two highlighted boxes indicate regions of interest. The video from the broadcast camera overlaid on the Omnicam video (note difference in perspective).

3. Method

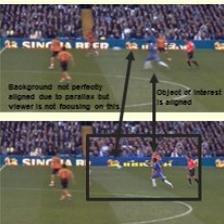
To find the region in the Omnicam that "best corresponds" to what the broadcast camera is capturing, taking into account the offset between the cameras:

1. Make a simple 3D model of stationary parts of the world, including the ground plane. Measure the positions of the cameras, and track the pan / tilt / zoom of the Broadcast Camera as it moves using line / edge / corner tracking.
2. Detect people in the broadcast camera image using motion-compensated background subtraction, and find the point in the image where the biggest person's feet touch the ground.
3. Calculate the 3D position of the reference point based on the assumption of contact with the ground plane and use this to define the depth in the scene that best represents the location of the broadcast camera image.
4. Create a new "Virtual Camera" from the viewpoint of the Omnicam, that covers the view of the broadcast camera at the depth of the reference point.



Reference point for depth of interest: assume lowest point of largest segmented person is contact with the ground plane.

4. Results



Omnicam image

- We want to give the viewer the option to make a smooth transition to the area of interest covered by a broadcast camera.

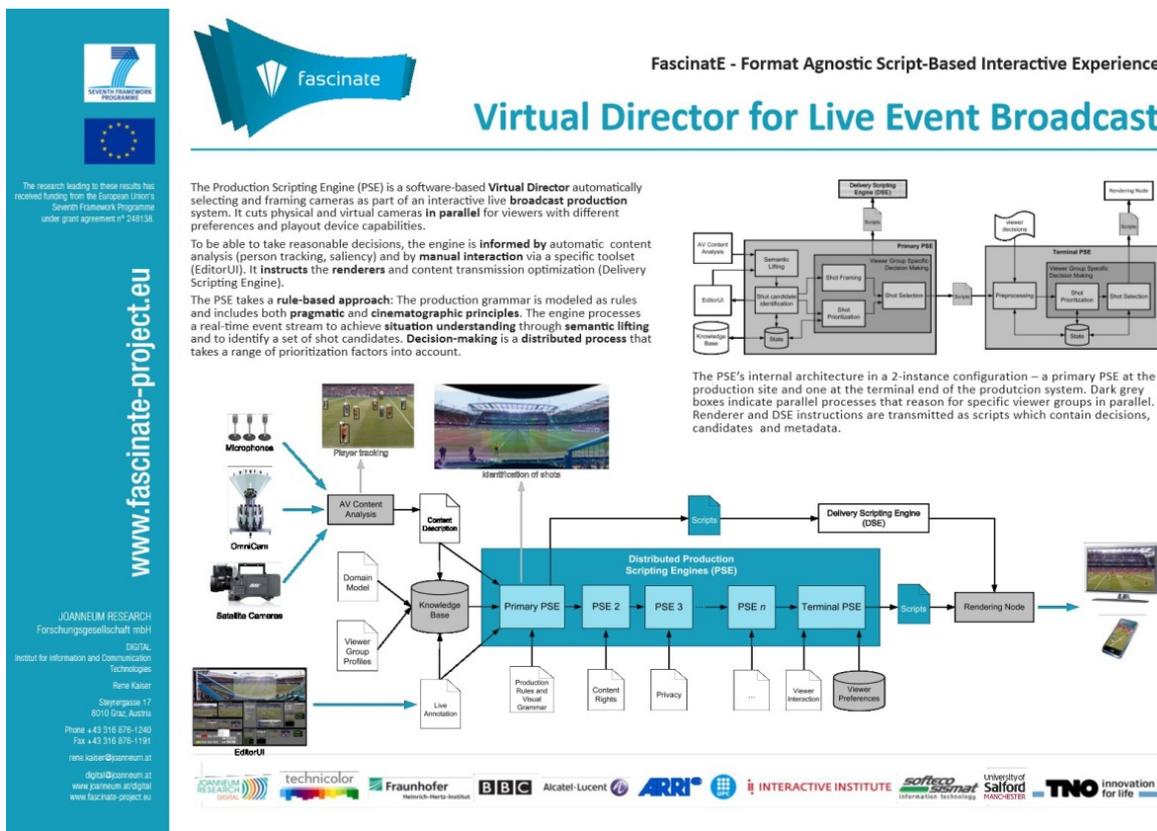
Overlaid broadcast camera image:

- Aligned to keep player in the same place rather than the background
- A button from the portion of the Omnicam image selected in this way to the broadcast camera image does not disturb the viewer, as there is no "jump" in the position of the object they are looking at.

Figure 12: Poster for broadcast camera integration demo

4.3 Virtual Director

The Virtual Director automatically frames and cuts between virtual camera shots. It takes personalized decisions in parallel for viewers with different preferences and end-user device capabilities. The engine is informed by automatic content analysis, through manual interaction via an Editor User Interface and a semantic data store. It instructs the renderer to change the viewpoint within the panoramic video, respecting cinematographic principles. The demo used recorded video to showcase prototypical production behaviour in two domains, soccer and dance performance.



4.4 Content Analysis

Through Content Analysis, tracks of people and information about visually-salient regions are automatically extracted from the panoramic video. The locations of people in the panorama serve as input for the Virtual Director and support its decision making process. The challenge here is to process the ultra-HD panorama in real-time. The saliency detection automatically extracts those regions of the panoramic video to which a human viewer's attention will be most likely be drawn. The algorithm provides potential regions of interest, which can be used for making editing decisions and to optimise content transmission.




The research leading to these results has received funding from the European Union's Seventh Framework Programme under grant agreement n° 348136.

www.fascinate-project.eu

JOANNEUM RESEARCH
Forschungsgesellschaft mbH
DIGITAL
Institut für Information and Communication
Technologies
Ralph Kasper
Steinringgasse 17
8010 Graz, Austria
Phone: +43 316 876-1240
Fax: +43 316 876-1191
georg.stralinger@joanneum.at
digital@joanneum.at
www.joanneum.at/digital
www.fascinate-project.eu

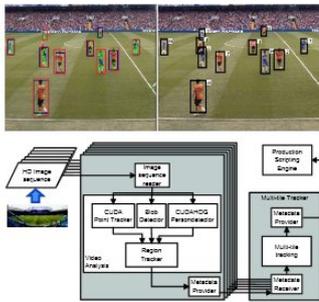


FascinatE

Automatic Content Analysis

Real-time Person Detection and Tracking in Ultra-HD Panoramic Video

- ◆ Real-time person tracking on 7Kx2K-Panorama Video Stream (85 ms/frame)
- ◆ **CUDA Point Tracker:**
Fermi architecture 5k-10k Points, 9 ms/frame
- ◆ **CUDA HOG Person Detector:**
fastHoG, scale ratio 1.3 70 ms/frame
- ◆ **Blob Detector:**
OpenCV, for motion blur 35 ms/frame
- ◆ **Region Tracker:**
Fusing of detected regions and feature points
Appearance/Motion model based 15 ms/frame
- ◆ **Multi-tile Tracker:**
Fusing of Results for stitched panoramic image



Automatic Detection of Salient Regions

- ◆ Visual salient regions extracted from panorama video stream
- ◆ Spatio-temporal saliency histogram for each image grid cell
- ◆ Centre-surround model from the difference of Gaussian image pyramid
- ◆ Pixel intensities and colour opponents as features
- ◆ Compare histograms of current frame with recent histograms in buffer
- ◆ Rectangular salient regions stabilized using Kalman filter

- ◆ Applications
- ◆ Support dynamic encoding of tiled panoramic content
- ◆ Region of interest for virtual camera viewport



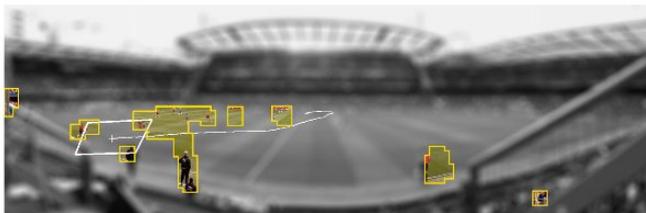




Figure 14: Poster for content analysis demo

4.5 Delivery using Adaption in the Network

Interactive UHDTV Delivery to any device was demonstrated using the scalable delivery system developed in FascinatE. It allows a broad range of end-user devices, from low-powered mobile devices to high-end immersive setups, to best use their bandwidth and processing resources for interactive access to any region of the panoramic video. The demonstration showed how intermediate network elements can adapt video streams on-the-fly and seamlessly combine efficient video broadcast with on-demand delivery. In addition, a Rendering Proxy demonstrated in-network rendering on behalf of multiple mobile terminals. Several tablets were available in the Digital Performance Lab for the audience to interactively view and navigate around the live panoramic video and previous FascinatE recordings. The 9 views produced by the Rendering Proxy were jointly displayed on a HD monitor. The behaviour of each tablet users in terms of navigation in the panoramic video has been recorded for further analysis.



Figure 15: Set-Up of the offline Interactive UHDTV Delivery set-up and HD monitor (left) - Recording of the 9 live interactive sessions for tablet, as produced by the Proxy Renderer (right)

The poster for the scalable delivery demo is titled "Scalable Delivery System for Interactive Ultra-HD Video" and features the Fascinate logo and Alcatel-Lucent branding. It is organized into several sections:

- Multi-Resolution Tiled Representation:** Shows an "Ingest" process where a video is split into tiles (e.g., 1.1, 1.2, 1.3, 1.4, 1.5, 1.6). These are then processed through "Dynamic Segmentation" and "H.264 encoding" to create "FascinatETiles". These tiles are described as "Independently Decodable Streams" that allow a client to "selectively access any region at any scale". It also mentions "Optimisation of Coding and Tiling Schemes" including Saliency-Based Rate Allocation, Multi-Quality and Resolution Layering, and analysis of tiling support in H.26x standards.
- PubSub Transport: Adaptive Mix of Unicast & Multicast:** Illustrates a "Seamless combination of Multicast: push tiles useful for many" and "Unicast: pull tiles on-demand". The flow goes from "Ingest" (Downsampling, Scene Segmentation, Encoding) through a "Relay" (Dynamic Transport Selection, HAS Relay, Temporal Segmentation) to "Proxy #1, #2, ..., #n". These proxies serve "Simulated Random Clients".
- Rendering Proxy: Interactive Panoramic Video to any Device:** Lists benefits: "In-Network rendering for legacy devices", "Thin client (Android tablet) sends navigation commands", "Scalable rendering by sharing GPU resources across clients", and "Achieves perfect random access". A diagram shows a "Segment Client" sending commands to a "Rendering Proxy" which then outputs to "Client #1, #2, ..., #n".
- Integrated Set-Up:** A detailed network diagram showing the physical and logical connections between servers, proxies, and clients.

At the bottom, the poster lists project partners: JOANNEUM RESEARCH, technicolor, Fraunhofer, BBC, Alcatel-Lucent, ARRI, INTERACTIVE INSTITUTE, softico, sismat, University of Salford, and TNO. It also includes contact information for the project website and several team members: Nico Verzijp, Patrice Rondao Aifonso, Jean-François Macq, Alcatel-Lucent, and Bell Labs, Antwerp, BE.

Figure 16: Poster for scalable delivery demo

4.6 Delivery using Tile-based HTTP Adaptive Streaming

This demonstration showed interactive delivery of UHDTV video to TVs and tablets using adaptive streaming technology. With tiled HTTP adaptive streaming (HAS), video is spatially segmented into a grid of video tiles. Each video tile is encoded separately and then temporally segmented according to common HAS solutions. Users could control a virtual camera with navigation commands on multiple tablets, select predefined regions-of-interest, or playback broadcast camera content on the tablet itself. Features included trick play, selecting predefined regions of interest and event signalling. For tablets, the iCOMBINE application offered navigation into panoramic video and interaction sharing. During the demonstrations we showed the TV-based solution with second screen control; additionally the audience could test and experiment with the iCOMBINE application on one of the eleven tablet devices, including the option to record interactions and replay them on another device.

The poster is titled "Interactive UHDTV in the Future Media Internet" and is sponsored by TNO innovation for life. It features the Fascinate logo and logos for the European Union and the Seventh Framework Programme. The poster is divided into several sections:

- Media Trends:** Lists four trends:
 - Increased realism through higher resolutions and larger displays (with images of large screens).
 - Increased quality of experience via adaptive streaming (with a cloud and server icon).
 - Increased personalisation via second screens (with a TV and tablet icon).
 - Increased service reach via mobile applications (with a smartphone icon).
- Tiled HTTP Adaptive Streaming:** A flowchart showing the process from a "Layered Scene" through "Layer BCam", "Scale MD", "Grid 4x4", "Tile T1-T4", "Representation B1-B2", and "Generic Segment" to "Time Segment".
- Second Screen Interaction:** A diagram showing the flow from a "Segment Server" (via "Tiled-HAS delivery" and "Event signalling") to a "Video Renderer" (containing "Segment Client", "Frame Combiner", and "Metadata Handler"), which then outputs to a "Display" and a "second screen tablet" (receiving "PTZ-commands" and "Predefined events").
- Mobile Applications: iCOMBINE:** Lists features:
 - FascinatE-as-an-App
 - Tiled adaptive streaming using HLS and DASH
 - Features script-based interaction sharing
 Includes images of the iCOMBINE app interface and a comparison of a wide-angle stadium view vs. a zoomed-in view.

At the bottom, there is a list of partner logos including JOHANNES KEPLER UNIVERSITÄT LINZ, technicolor, Fraunhofer, BBC, Alcatel-Lucent, ARRI, INTERACTIVE INSTITUTE, softco, Sismat, University of Salford, and TNO innovation for life.

Figure 17: Poster for Adaptive UHDTV Streaming



Figure 18: A user controls the virtual camera on a second screen tablet device.

4.7 Gesture Control

The Interactive Gesture Control explored new interaction possibilities within UHDTV content. It allowed the user to control the system using only their hands. The user was able to freely navigate the video by panning, tilting and zooming or to select among pre-defined region-of-interests (views). Furthermore, the user could perform natural and intuitive gestures to directly play or pause the video, decrease or increase the audio volume or mute it completely. Additionally, the gesture control system identified and recognised faces allowing the system to create a hierarchy where several options and actions are only available to specific users.



The research leading to these results has received funding from the European Union's Seventh Framework Programme under grant agreement n° 248138

www.fascinate-project.eu

JOHANNEM RESEARCH
Forschungsgesellschaft mbH
DIGITAL – Institute for Information and
Communication Technologies
Steyrigasse 17
A-8010 Graz
Austria
Tel: +43 316 876 1240
Fax: +43 316 876 1191
grip@johannem-research.com
www.johannem-digital.com
www.fascinate-project.eu



FascinatE – Format Agnostic Script-Based Interactive Experience

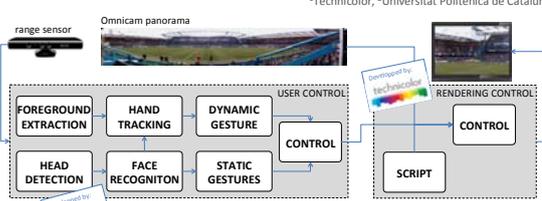
FascinatE Rendering Node with ROI zoom and Gesture Control

CONTRIBUTION

Real-time hands tracking and gesture recognition using depth information to create a user interface based on a 3D virtual tablet with gesture shortcuts.

Script controlled real time rendering platform to allow **Virtual Camera Navigation** of region of interest with pan, tilt, zoom.

Matle Borsum¹, Axel Kochale¹, Albert Gil², Javier Ruiz-Hidalgo³ and Xavier Suau²
{matle.borsum, axel.kochale}@technicolor.com {albert.gil, j.ruiz, xavier.suau}@upc.edu
¹Technicolor, ²Universitat Politècnica de Catalunya



USER CONTROL TECHNOLOGIES

- Static Gestures:** Detects and classifies a customizable set of static and natural gestures.
 - Full scan: control request by any user
 - Focused scan: current user command
 - Custom gesture: The system is able to learn new gestures customized to each user
- Head Tracking:** Detects and tracks human heads (users). Initialized by a 'take control' gesture.
- Hands Tracking:** One/Two Open/Closed hands tracked in a 3D virtual tablet located in front of the tracked head.
- Dynamic Gestures:** Detects and classifies hands movements, such as swiping or circling.

NAVIGATION

Menu & View Selection / Panoramic Navigation

Users can select any option of the menu by using the 'L' gesture. They can also browse through the different views (ROIs). They can navigate through the panorama by moving the 'L' gesture in a 3D space.

Custom

Users can customize their own gesture to show the complete panoramic scene.

GESTURE SHORTCUTS

Take Control

The 'Tea' gesture is used as 'Start/Home' button.

Volume Down/Up & Mute

Users can decrease the volume (left image), increase it (centre image) or mute/unmute the audio by the 'cross' gesture (right image).

Pause

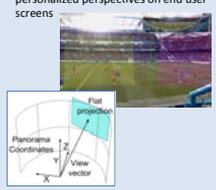
Users can pause and resume the player by simulating the 'pause icon' with their hands.

SCRIPT-CONTROLLED RENDERING

XML script controls visual rendering:

- Camera cluster offers multiple region of interest (panorama, shot cameras)
- Live capture requires automation of workflow

Rendering converts *Layered Scene* into personalized perspectives on end user screens



FORMAT-AGNOSTIC CONTENT PRODUCTION

Variety of end terminals require **format agnostic production**

Terminal and services supply **personalized visual perspective**

Content interaction beyond channel swapping: Pan, Tilt, Zoom

Scalable platform architecture allows application on different target terminals



CONCLUSIONS

Real time capable terminal **platform** for **pan, tilt, zoom** navigation within a panoramic scene by selecting script options such as region of interest updates

Easy personalization by multiple **depth processing technologies** in **parallel** to create a **3D virtual tablet** in front of a user and to recognize **customizable natural gestures** for specific (shortcut) commands.











Figure 19: Poster for gesture control and video rendering demo



Figure 20: Setup for the gesture control stand-alone demonstration

4.8 Video Rendering

Video Rendering was demonstrated in multiple settings. The laptop-based demos showcased format-agnostic production services allowing navigation within live content for a customised perspective experience using a mouse or trackball. The user could select the individual perspective within ultra-high resolution content by panning, tilting or zooming the rendered view. Additionally, predefined regions-of-Interest were available to select a different perspective. This technology can be used as tool to prepare content for broadcast views like a virtual camera and allow end-users to act as their own director. The renderer was demonstrated in conjunction with the gesture control system (see poster in Figure 19).

4.9 3D Audio

FascinatE 3D audio uses an agnostic sound field description, enhanced with multiple audio objects, instead of using separate loudspeaker channels, which would require a specific loudspeaker layout. Higher Order Ambisonic (HOA) recording is carried out using a spherical microphone array consisting of 32 capsules. These signals are recorded, encoded into the HOA format, transported and mixed with additional HOA encoded audio objects. The HOA format can be decoded to any existing loudspeaker setup. Remixing for new loudspeaker setups is not required and misplaced loudspeakers do not disturb the listening pleasure. Three demonstrations were shown of audio object extraction and rendering to both HOA and wave field synthesis systems illustrating adaptive rendering to a chosen viewpoint.

4.9.1 Audio Object Extractor

The first demo was a video of an example implementation of the audio object extractor for deriving the on-pitch sounds for a football broadcast application. The accompanying poster described the purpose of the audio object extractor, contextualising it within the FascinatE audio system (as shown in Figure 21).

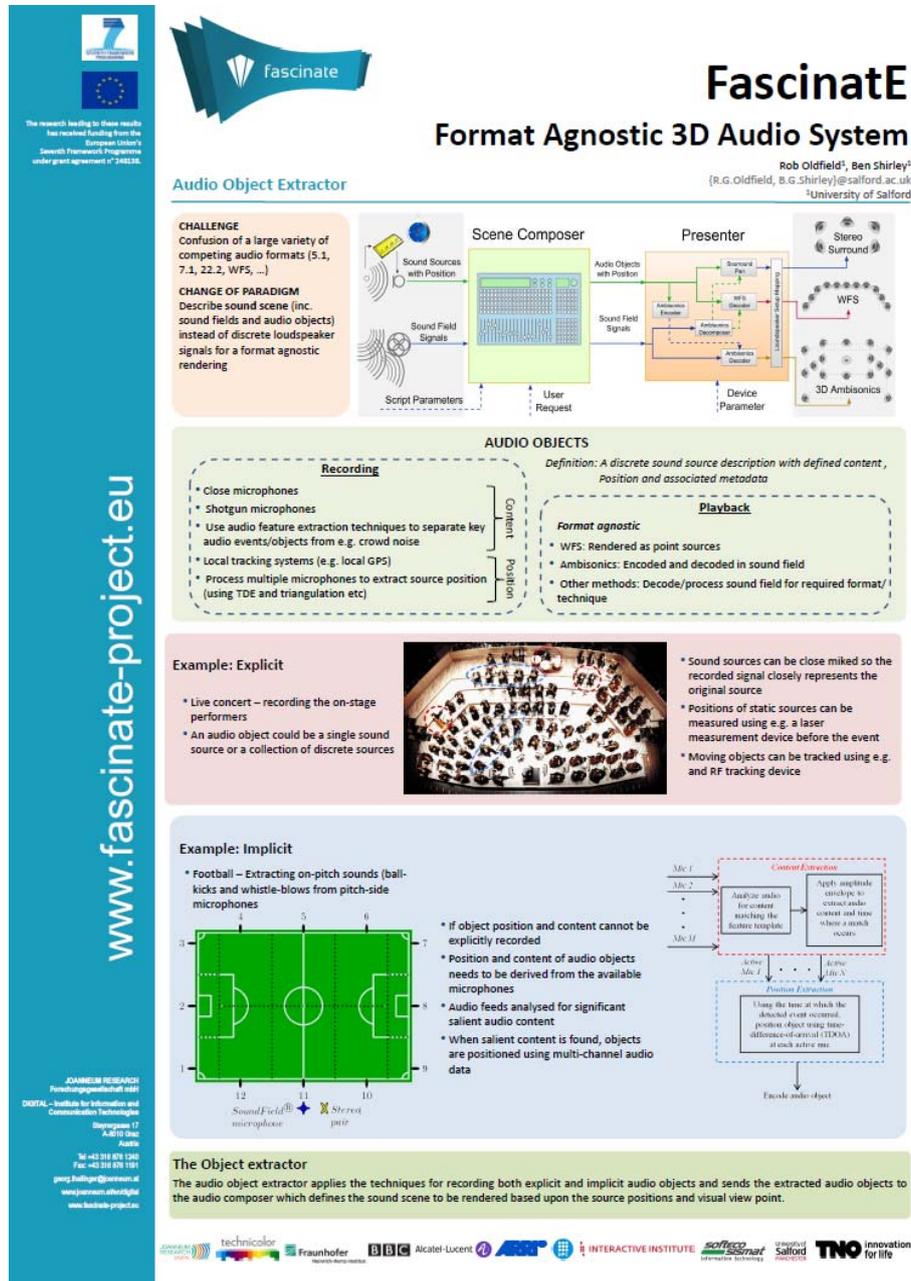


Figure 21: Audio object extraction demo poster

4.9.2 Higher Order Ambisonics

The second audio demo showed the 3D Higher Order Ambisonics audio with adaptive rendering with user interaction. The audio presenter rendered several sound fields recorded with the Eigenmike during the previous FascinatE test shoots over the 17 channel higher order Ambisonics system. As the user panned in the visual content, the sound field rotated to match to give an interactive experience.

The research leading to these results has received funding from the European Union's Seventh Framework Programme under grant agreement n° 246135.

JOHANNES KEPLER UNIVERSITÄT LINZ
Forschungsinstitut für
Digitale – Institute for Information and
Communication Technologies
Inneregasse 17
A-4020 Linz
Austria
Tel: +43 316 876 1240
Fax: +43 316 876 1393
prog.fhdg@jku.ac.at
www.jku.ac.at/fhdg
www.fascinate-project.eu

FascinatE

Format Agnostic 3D Audio System

FascinatE Audio Rendering Node for Real 3D Sound with Height

Holger Kropp¹, Jens Spille¹, Rob Oldfield², Ben Shirley²
(Holger.Kropp, Jens.Spille)@technicolor.com (R.G.Oldfield, B.G.Shirley)@salford.ac.uk
¹Technicolor, ²University of Salford

Today: confusion by variety of existing audio formats
5.1, 7.2, 22.2, WFS, ...

Next: change of paradigm
Describe sound field instead of discrete loudspeaker signals

Recording

- Higher Order Ambisonics (HOA) requires microphone arrays and encoding
- Audio Objects are encoded to HOA

Playback

- Agnostic to the loudspeaker setup "one format fits all"
- Superior spatial resolution up to real 3D sound with height
- Dynamically modified viewpoints
- Sound field rendering systems HOA

Football

Microphone setup used for Chelsea test shoot

Dance Performance

Microphone setup used for Carmen for test shoot

Recording

- Eigenmike with 32 capsules
- Ambisonics encoder matrix
- 4th order Higher Order Ambisonics (HOA)

Playback

- Ambisonics decoder matrix
- 4th order HOA
- Rendering onto real setup 18 loudspeakers
- Headphones using Head Related Impulse Responses

SOUND FIELD

Higher Order Ambisonics Ext.-Wave / B-Format

CONCLUSIONS

Format agnostic full 3D sound (incl. height) recording, transmission and playback is achieved. Adaption to individual loudspeaker setups.

Figure 22: Poster for 3D audio rendering demo

© FascinatE consortium: all rights reserved

page 19

4.9.3 Wavefield Synthesis

The third offline audio demonstration showed interactive audio rendering over a 64 channel wave field synthesis system (Figure 23). The users could pan and zoom into the visual panoramic content from the Proms test shoot with the audio content adapting accordingly.



Figure 23: Wave field synthesis audio system

The research leading to these results has received funding from the European Union's Seventh Framework Programme under grant agreement 258255.

FascinatE

Format Agnostic 3D Audio System

FascinatE Audio Rendering Node for Volumetric Audio

CHALLENGE
Confusion of a large variety of competing audio formats (5.1, 7.1, 22.2, WFS, ...)

CHANGE OF PARADIGM
Describe sound scene (inc. sound fields and audio objects) instead of discrete loudspeaker signals for a format agnostic rendering

Recording

- Higher Order Ambisonics (HOA) requires microphone arrays and encoding
- Audio Objects are recorded using various techniques and can be encoded to a HOA representation

Playback

- Agnostic to the loudspeaker setup "one format fits all"
- Superior spatial resolution up to real 3D sound with height
- Dynamically modified viewpoints
- Sound field rendering systems HOA

Football

Microphone setup used for Chelsea test shoot

Dance Performance

Microphone setup used for Carmen for test shoot

Wave Field Synthesis

- Multiple loudspeakers recreate physical wave fronts
- Volumetric solution in 2D with depth
- No sweet spot
- Audio objects reproduced as point sources
- Sound fields reproduced using plane wave decomposition or hybrid techniques

CONCLUSIONS

Format agnostic full 2D (with depth) recording, transmission and playback is achieved for reproduction in a space with no sweet spot. Adaptive sound to match viewpoint including pan, zoom and rotation of audio objects and sound fields.

JOANNEUM RESEARCH
Forschungsinstitut für
DIGITAL – Institute for Information and
Communication Technology

Heinrichstraße 17
42699 Solingen
Aachen

tel: +49 212 878 1240
fax: +49 212 878 1191
perry.kollmann@joanneum.at
www.joanneum.at/foi/foi/foi

Figure 24: Poster wave field synthesis audio demo

5 Coverage in the Press

The demonstrations were written about by several publications, in addition to reports and blogs produced by the consortium members themselves.

5.1.1 Book of the Future

Book of the Future (on-line publication about how technology may affect the future, Figure 25) <http://www.bookofthefuture.co.uk/2013/05/fascinate-the-future-of-event-television/#.UaiN96Lvskp>

The screenshot shows the website 'BOOK OF THE FUTURE' with a navigation menu (Live, Content, Services, About) and a search bar. A large blue graphic with the word 'fascinate' is prominent. The main article is titled 'Be The Director: The Future of Event Television', posted on May 31, 2013, by Tom Cheesewright. The article text discusses the shift from passive television to interactive experiences. A sidebar on the right features 'VIEWS ON THE FUTURE' and 'Recent Posts' including 'Why Electronics Stores Are as Dead as the DVD'. A 'Tweets' section on the right shows several tweets related to the article and the website.

Figure 25: Coverage at bookofthefuture.com

5.1.2 One Publication

One publication (On-line publication featuring the latest news and opinions about professional broadcast applications and technology worldwide, Figure 26) <http://www.one-publication.com/2013/05/30/multi-million-euro-fascinatE-interactive-panoramic-ultra-hd-system-demonstrates-in-uk/>

The screenshot shows the homepage of 'ONE PUBLICATION'. The main headline reads: 'Multi-million euro FascinatE interactive panoramic Ultra HD system demonstrates in UK'. The article is dated May 30, 2013, by Chris Dickinson. Below the headline is a large photograph of a stadium with a camera rig on a platform. The rig includes a checkered target and long poles. A person in a red jacket is visible near the rig. The stadium seating is blue and yellow. The website has a navigation menu with 'Home', 'Articles', 'Videos', 'About us', 'Subscribe', 'Download', and 'Contact us'. On the right side, there are sections for 'FOLLOW US!' with social media icons, a search bar, a 'SIGN UP' form for email, and a 'TELL US YOUR STORY' section with a megaphone icon. At the bottom right, there is a 'LATEST TWEETS' section featuring a tweet from HarrisBroadcast.

Figure 26: Coverage at one-publication.com

5.1.3 Manchester Gazette

The Manchester Gazette (a local newspaper for the Manchester area, Figure 27) <http://manchestergazette.co.uk/salford-university-brings-the-future-of-live-tv-to-mediacityuk/>

Manchester Gazette | Wednesday, June 19, 2013

The Manchester Gazette

Home Business Crime Education Entertainment Local Politics Sport Technology

You are here: Home » Featured » Salford University brings the future of live TV to MediaCityUK

Salford University brings the future of live TV to MediaCityUK

Posted: 10:24 AM, June 7, 2013 by Chief Editor



The future of live television, where the viewer becomes the director, has been unveiled in the cutting-edge surroundings of the University of Salford's MediaCityUK campus.

University researchers were joined by partners from 10 technology companies, broadcasters, higher education institutions and other organisations across Europe for the final demonstration of the FascinatE Project – an £8m EU-funded study examining the next generation of interactive broadcast media technology.

- Man arrested following rape of 13-year-old in Wigan**
Police investigating the rape of a 13-year-old girl in ...
FULL ARTICLE
- Stagecoach Manchester Staff Answer the Charity Call for St Ann's Hospice**
A group of kind-hearted customer service switchboard operators have ...
- THE NEW HOME OF OLDHAM BOXING AND PERSONAL DEVELOPMENT CLUB IS A KNOCK OUT**
Oldham Boxing and Personal Development Club (OBPDC) is proud ...
- Police Appeal Following Gang Rape of Woman at Gorton Party**
Police are investigating after a woman was raped by ...

Figure 27: Coverage in the Manchester Gazette

6 Reflections on the Demonstration

This demonstration allowed all of the key elements of the project to be shown in an integrated live end-to-end system. These included:

- A new more compact version of the OmniCam, capable of 360-degree operation
- Real-time tracking of a satellite HD camera with its viewing region being highlighted in the panoramic image and its video being streamed to the terminal
- The final version of the gesture-based interaction system
- Rotation and zooming of the audio sound field and audio objects to match the viewing direction as the user navigated around the panorama
- The final version of the Production Scripting Engine, providing automated selection of regions-of-interest from within the panoramic image (a kind of “virtual director”)
- Live delivery chain, including real-time interfaces (raw panoramic video) with the OmniCam and the Video Terminal (FRN), dynamic combinations of unicast/broadcast transport, in-network rendering for multiple thin clients, and real-time creation of video segments for HTTP adaptive delivery.

6.1.1 Visitor Feedback

Visitors that registered online were asked to complete a survey after the event. The response rate was about 20% (12 responses out of 60 places available in the main demo). The responses indicated that the majority of visitors rated the event as “Excellent” (Figure 28).

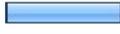
Overall, how would you rate the event?			
		Response Percent	Response Count
Excellent		66.7%	8
Very good		8.3%	1
Fairly good		25.0%	3
Not good at all		0.0%	0
answered question			12

Figure 28: Headline results from the visitor survey

There were a few technical glitches during the live demos, for example one of the PCs running the audio part of the demo crashed just before one performance, and region-of-interest data from the broadcast camera was routed to the wrong rendering PC during another. Nevertheless, the project fully achieved its aim of showing a complete live end-to-end chain, proving the feasibility of implementing all key parts in real time.

Some visitors commented that it was very ambitious to attempt to show all these elements of the project operating together during a live performance, in such a way that each could be explained to visitors within the time available. The visitor comments on the live demo received via the online survey are shown in Figure 29.

Q3. Comments on the streamed broadcast session in the Egg Suite:	
1	I was late, I ended up in the DPA
2	Very good
3	Very successful, though it could have done with a bigger space and the projector screen was too low.
4	Inspiring and a privilege to view
5	Very innovative (and brave!) to be streaming it live and to have full live demos of the various aspects of the system]
6	Very good to see the live demonstration; it started a bit late though, and the speeches went on for a bit too long
7	Amazing technological showcase, slightly marred by the technical issues

Figure 29: Visitor comments on the live demo

6.1.2 Further Technical Details of the Demonstrations

Technical details of the FascinatE system as presented at this demo will be documented in [D1.5.3] (project internal), and the performance of the system will be summarised in [D1.5.4] (publicly-available from the end of July 2013). Evaluations carried out at the final demo are reported in the project-internal deliverables [D5.1.4] (gesture-based navigation and tablet interaction) and [D1.1.3] (virtual director evaluation).

6.1.3 Acknowledgements

The project is very grateful to the composer Stephen Davismoon, the band Bears?Bears! and the dancers and vocalists involved in the production *Deeper than all Roses*, for their permission to use the performance as the centrepiece for the live demonstration, and their flexibility in accommodating the requirements of the project.

The project also thanks Jan Badley, Mike Polshaw and the rest of the event and facility organisers at the University of Salford, for being so helpful and efficient in helping the project stage its final demonstration.



7 References

- [D1.1.3] FascinatE internal deliverable D1.1.3 “Final requirements document” (not publicly available).
- [D1.5.3] FascinatE internal deliverable D1.5.3 “Final System Integration” (not publicly available).
- [D1.5.4] FascinatE public deliverable D1.5.4 “Report on system performance”, available via the project website <http://www.fascinate-project.eu/index.php/publications/public-deliverables/> at the end of July 2013.
- [D5.1.4] FascinatE internal deliverable D5.1.4 “Real-time AV renderer with support for WFS and full interactivity” (not publicly available).
- [D6.1.1] FascinatE public deliverable D6.1.1 “First Demonstration”, available via the project website <http://www.fascinate-project.eu/index.php/publications/public-deliverables/>
- [D6.1.2] FascinatE public deliverable D6.1.2 “Interim Demonstration”, available via the project website <http://www.fascinate-project.eu/index.php/publications/public-deliverables/>

8 Glossary

Terms used within the FascinatE project, sorted alphabetically.

FascinatE	Format-Agnostic Script-Based Interactive Experience
IBC	International Broadcast Convention

Partner Acronyms

ALU	Alcatel-Lucent Bell NV, BE
ARI	Arnold & Richter Cine Technik GMBH & Co Betriebs KG, DE
BBC	British Broadcasting Corporation
DTO	Technicolor, DE
HHI	Heinrich Hertz Institut, Fraunhofer Gesellschaft zur Förderung der Angewandten Forschung e.V., DE
JRS	JOANNEUM RESEARCH Forschungsgesellschaft mbH, AT
SES	Softeco Sismat S.P.A., IT
TII	The Interactive Institute, SE
TNO	Nederlandse Organisatie voor Toegapast Natuurwetenschappelijk Onderzoek – TNO, NL
UOS	The University of Salford, UK
UPC	Universitat Politecnica de Catalunya, ES

9 Appendix 1: Flyer for Event



The research leading to these results has received funding from the European Union's Seventh Framework Programme under grant agreement n° 248138.



FascinatE –

Format Agnostic SScript-based INterActive Experience

www.fascinate-project.eu

Project Overview

The FascinatE project has developed a system to allow end-users to interactively view and navigate around an ultra-high resolution video panorama showing a live event, with the accompanying audio automatically changing to match the selected view. Using this system, the output is adapted to the specific end-user device, covering anything from a mobile handset to an immersive panoramic display. At the production side, new audio and video capture systems have been developed, as well as scripting systems to control the shot framing options presented to the viewer. Intelligent networks with processing components have been built to repurpose the content to suit different device types and framing selections. End-user devices now support innovative interaction methods to allow viewers to navigate and display the content.

Main Innovations

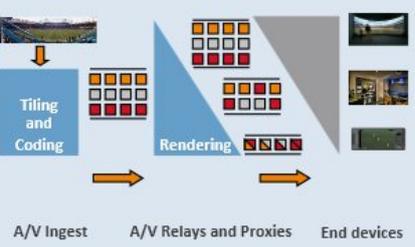
Layered Panoramic & Omnidirectional A/V Capturing and Scene Production



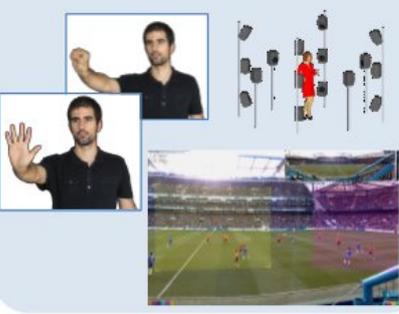
Production Automation



Scalable and Interactive Delivery



Immersive and Interactive Applications for End-Users







The research leading to these results has received funding from the European Union's Seventh Framework Programme under grant agreement n° 248138.

www.fascinate-project.eu

JOANNEUM RESEARCH
Forschungsgesellschaft mbH

DIGITAL – Institute for
Information and
Communication Technologies

Steyrergasse 17
A-8010 Graz
Austria

Tel +43 316 876 1240
Fax: +43 316 876 1191
georg.thallinger@joanneum.at
www.joanneum.at/en/digital
www.fascinate-project.eu

#FascinatEdemo



FascinatE –

Format Agnostic SCRIPT-based INTERACTIVE Experience

Facts and Figures

FascinatE is a €9.35m EU-funded project involving a group of 11 partners from across Europe, running from 1st February 2010 to 31st July 2013. FascinatE stands for *Format-Agnostic SCRIPT-based INTERACTIVE Experience* and is looking at broadcasting live events to give the viewer a more interactive experience no matter what device they are using to view the broadcast.



Fascinate_Prjct



FascinatE.FP7



FascinatE

You can find out more on our website, or follow us via Twitter, Facebook and Vimeo.

Projects Partners



10 Appendix 2: Event Guide



The research leading to these results has received funding from the European Union's Seventh Framework Programme under grant agreement n° 248138.

www.fascinate-project.eu

JOANNEUM RESEARCH
Forschungsgesellschaft mbH

DIGITAL – Institute for
Information and
Communication Technologies

Steyrergasse 17
A-8010 Graz
Austria

Tel: +43 316 876 1240
Fax: +43 316 876 1191
georg.thallinger@joanneum.at
www.joanneum.at/en/digital
www.fascinate-project.eu

#FascinatEdemo



FascinatE – Showcase: Future Broadcast Technology

Final Demonstration of the EU FP7 FascinatE Project

The final demonstration of FascinatE features the premiere performance of '**deeper than all roses**', the latest large-scale music composition from Stephen Davismoon, featuring band **Bears?Bears!** Zoe Swindells, Nick Rogers, Sam Jones, Tom Miller, and Benjamin Gow with guest vocalist Anikó Tóth, and live performance artists Joseph Lau and Shona Roberts, in celebration of the works of the celebrated American poet E.E. Cummings.

Various innovative aspects of the FascinatE system will be demonstrated during the live performance, shown three times during the day. The performance and acquisition space will be open during all performances and offline demonstrations will be held throughout the day.

Programme: Live Demonstration

- ❖ Project Overview and Welcome Mr. Georg Thallinger,
Joanneum Research
- ❖ Showcase Introduction Prof. Graham Thomas, *BBC*

- ❖ Phase 1 – Totally Immersive Experience Prof. Graham Thomas, *BBC*
 - ❖ Live panoramic UHDTV
 - ❖ Immersive 3D audio
 - ❖ Gesture-based scene navigation
 - ❖ Integrated broadcast camera coverage
- ❖ Phase 2 – Interactive UHDTV Meets Network
 - ❖ Live UHDTV transmission
 - ❖ Multi-screen display and interaction
- ❖ Phase 3 – My Personalised Production
 - ❖ Virtual Director
 - ❖ Automatic content analysis
 - ❖ Manual annotation interfaces
- ❖ Questions and wrap-up Mr. Georg Thallinger,
Joanneum Research



The research leading to these results has received funding from the European Union's Seventh Framework Programme under grant agreement n° 248138.

www.fascinate-project.eu

JOANNEUM RESEARCH
Forschungsgesellschaft mbH

DIGITAL – Institute for
Information and
Communication Technologies

Steyrergasse 17
A-8010 Graz
Austria

Tel +43 316 876 1240
Fax: +43 316 876 1191
georg.thallinger@joanneum.at
www.joanneum.at/en/digital
www.fascinate-project.eu

#FascinatEdemo



Demonstrations and Technologies

The **OmniCam** is a 180° panoramic video acquisition system. For the live demonstration, a new version of the omnidirectional camera is used. This new OmniCam has a much smaller form factor compared to the first version, that was equipped with the Alexa M by ARRI. This new OmniCam is equipped with 10 Indie GS2K cameras mounted on a mirror rig with 10 mirrors capable to capture a 360° cylindrical field of view. For the live demo, the video output from five cameras is used to create 180° panoramic video. The mirror rig with the cameras has a height of 60cm and a diameter of 50cm. The weight is about 15 kilograms. The acquisition is performed at a frame rate of 25 frames/second.

Broadcast Camera integration. Live insertion of broadcast camera video on top of the OmniCam panorama. This requires tracking the pose of the camera in real-time, estimation of the object of interest within the scene, and integration with a 3D model of the scene. Given this data the OmniCam determines which region most closely matches the view from the broadcast camera. A renderer program is demonstrated which can superimpose the broadcast camera video over the panorama accordingly.

The **Virtual Director** automatically frames and cuts virtual camera shots. It takes personalised decisions in parallel for viewers with different preferences and end-user device capabilities. The engine is informed by automatic content analysis, through manual interaction via an Editor User Interface and a semantic data store. It instructs the renderer to change the viewpoint within the panoramic video, respecting cinematographic principles. This demo uses recorded video to showcase prototypical production behaviour in two domains, football and dance performance.

Through **Content Analysis**, person tracks and information about visually salient regions are automatically extracted from the panoramic video. The locations of persons in the panorama serve as input for the Virtual Director and support its decision making process. The challenge here is to process the ultra-HD panorama in real-time. The saliency detection automatically extracts those regions of the panoramic video that a human viewer's attention will be most likely drawn to. The algorithm provides potential regions of interest, which can be used for making editing decisions and to optimise content transmission.

Interactive UHDTV Delivery to Any Device is demonstrated using the scalable delivery system developed in FascinatE. It allows a broad range of end-user devices, from low-powered mobile devices to high-end immersive setups, to best use their bandwidth and processing resources for interactive access to any region of the panoramic video. The demonstration shows how intermediate network elements can adapt video streams on-the-fly and seamlessly combine efficient video broadcast with on-demand delivery. In addition, a Rendering Proxy demonstrates in-network rendering on behalf of multiple mobile terminals. Several tablets are available in the Digital Performance Lab for the audience to interactively view and navigate around the live panoramic video and previous FascinatE recordings.

Adaptive UHDTV Streaming demonstrates interactive delivery of UHDTV video to TVs and tablets. With tiled HTTP adaptive streaming (HAS), video is spatially segmented into a grid of video tiles. Each video tile is encoded separately and then temporally segmented according to common HAS solutions. Multiple users can control their own virtual camera with navigation commands on tablets, select predefined regions-of-interest, or playback broadcast camera content on the tablet itself, use trick play functionality and get signalling for events. For tablets, the *iCOMBINE* application offers navigation into panoramic video and interaction sharing.

Video Rendering is demonstrated in multiple settings. The laptop based demos showcase format agnostic production services allowing navigation within live content for a personalised experience using a mouse or trackball. The user can select the individual perspective within ultra-high resolution content by panning, tilting or zooming the rendered view, controlling their own virtual camera. Additionally, predefined regions-of-interest are available to instantly switch to a different viewpoint. This technology can be used as a tool to prepare content and allow end-users to act as their own director.

The **Interactive Gesture Control** explores new interaction possibilities within UHDTV content. It allows the user to control the system using only their hands. The user is able to freely navigate the video by panning, tilting and zooming or to select among pre-defined region-of-interests (views). Furthermore, the user can perform natural and intuitive gestures to directly play or pause the video, decrease or increase the audio volume or mute it completely. Additionally, the gesture control system identifies and recognises faces allowing the system to create a hierarchy where several options and actions are only available to specific users.

FascinatE 3D audio uses an agnostic sound field description, enhanced with multiple audio objects, instead of using separate loudspeaker channels, which would require a specific loudspeaker layout. Higher Order Ambisonic (HOA) recording is done with a spherical microphone array consisting of 32 capsules. These signals are recorded, encoded into the HOA format, transported and mixed with additional HOA encoded audio objects. The HOA format can be decoded to any loudspeaker setup. Remixing for new loudspeaker setups is not required and misplaced loudspeakers do not disturb the listening pleasure. Three demonstrations will be shown of audio object extraction and rendering to both HOA and wave field synthesis systems illustrating adaptive rendering to a chosen viewpoint.

