Panoramic Video: Design Challenges and Implications for Content Interaction

Goranka Zoric, Louise Barkhuus, Arvid Engström
Interactive Institute
Box 1197
SE-164 26 Kista, Sweden
{ goranka.zoric, louise.barkhuus, arvide }@tii.se

Elin Önnevall
Mobile Life @ Stockholm University
c/o SICS, Box 1263
SE-164 29 Kista, Sweden
elin@mobilelifecentre.org

ABSTRACT
In this paper we explore viewing and interaction in an emerging type of interactive TV, where viewers are presented with *panoramic ultrahigh-definition video* combined with *extensive interactive control* over view selection. Instead of delivering only what will be consumed, emerging TV services offer high-resolution panoramic video to the viewers, enabling them to more freely explore the broadcast content by selecting regions of interest and navigating within the larger panoramic image. However, as we open up the television space both in field of view and in terms of the freedom given to viewers, new interactional challenges emerge. We have done user studies on two systems for interacting with panoramic high-resolution video, one based on the tablet interaction and other on the gesture interaction. Our findings revealed a number of design challenges concerning properties specific to panoramic video. Based on findings from the user studies and the identified design challenges, we have compiled a set of the design recommendations on how to support interactive viewing of panoramic content.

Categories and Subject Descriptors
H.4 [Information Systems Applications]: Miscellaneous; H.5 [Information interfaces and presentation (HCI)]: User interfaces

General Terms: Design, Human Factors.

Keywords
Panorama, video, interactive TV, mobile TV, live sports, broadcasting, tablet, second screen, gesture interaction

1. INTRODUCTION
Within the area of interactive television there is a new emerging genre combining rich, panoramic live video with extensive interactive control over what to view within the covered scene. Extensive technical development in image capture and interaction modalities makes these high-resolution interactive TV images possible.

The services in this new genre build on current interactive television in two significant ways. First, they invite the viewer to *look into* a panoramic image space, as opposed to *looking at* framed images of a scene. Second, and closely linked to the above, they aim to provide the viewer with *production-like tools* for looking, i.e. for managing this increased control of view selection. These tools by necessity go beyond typical system controls such as audio level, colour and screen settings, since they extend the control into the image content itself. We are thus making the distinction between regular *system interaction* and what we may call *content interaction* – looking into the panoramic image space, using production-like tools for navigating and interacting directly with the image content as the broadcast or program unfolds.

Increased resolution and image size introduce numerous possibilities for viewer interaction as well as for integrating several screens into the viewing experience. With real-time content interaction, television viewers are slowly developing new ways of consuming shows, particularly live events. A sports match may be viewed on television while details are being looked up on a tablet device, or that device might even be showing a different version of the event. Individual viewers may immerse themselves in details of their choice within the broadcast, together or separately on "second screen" devices. At the same time, larger screens in the home serve as enablers for panoramic images.

In the last decade, increased attention has been paid to interactive TV (iTV), leading to development of novel systems offering advanced viewing features [1]. Consequently, a range of interactive functionalities has been suggested for viewers enabling more possibilities than those of standard remote controls. Much work is describing these technical approaches, where there are fewer studies that address how viewers would actually interact with content offering a multitude of possible views.

We here focus on *content interaction,* the possible interactivity that enables the viewers to compose a personalized television experience for themselves. We explore two main strains of interaction as well as look into more subjective perspectives from potential users. For the purpose of exploring these issues, we present three studies; a set of focus group studies conducted to explore basic potential interaction practices among television viewers, and two user studies of two different methods of interacting with the content: using a touch screen on a "second screen" tablet and using wide hand gestures for interacting with a large screen. In both cases interaction with panoramic video includes zooming in and out, panning and tilting, for navigating in the image space. The results of these studies lead to first a set of design challenges, which we subsequently frame as more specific design implications.

EuroITV'13, June 24–26, 2013, Como, Italy.

Figure 1: Example of a panoramic scene used

2. BACKGROUND AND RELATED WORK

We start this section by providing general observations on qualities of a panoramic image, and then concentrate on state of the art work focusing on systems enabling viewers to control a view selection and experiences around such. The two sections represent the two features that distinguish the new interactive TV genre we investigate; *panoramic view* and *real-time content interaction.*

2.1 Panoramic View

Why do we appreciate panoramic views? On a more general level Lansford and Jones [2] argue that there are values incorporated in a panoramic or a scenic view, which are the reason why properties with panoramic views, e.g. waterfronts are being highly valued. Meitner [3] suggests that a scenic beauty is "*determined jointly by the location of the viewpoint (observer) and the features of the view shed that is represented (experienced). In other words, scenic beauty ratings require a location/place, a landscape to appreciate from that location and a person to do the appreciating*". The experience of the panorama is thus dependent on the viewers' location. In the case of panoramic TV, deciding on what to look at is in this sense just as important as the "*location of the viewpoint*". Latour and Hermant [5] in their text Paris: Invisible City, are critical of what they see as the panorama's claims on capturing the entirety of a scene "*at at glance*", and suggest that "*it's time we updated our panoramas*". In their view, the panoramic view is too constructed to be a useful representation, in that it is dependent on a fixed viewpoint where the illusion is mastered. The above quotes all highlight the importance of the spectator's point of view in watching a panorama. In mediated panoramic events, such as the ones seen in the type of television investigated here, selecting a good point of view is an important consideration.

Wide formats for television and film in the home have seen a rise with the proliferation of large displays, but have their roots in cinematic formats. When the movie format Cinemascope became popular in Hollywood there were concerns with the width of the images [4]: "(…) *Hollywood's creative personnel feared that the wide screen would immobilize the camera and lead to long takes. Some editors were afraid to cut quickly, worrying* ***that viewers would not know where to look*** *in a rapid series of wide compositions*" (our bold). These concerns are brought to the fore again with increased interactive control. However, the action and content in the scene play an important role in guiding the viewer. For example, if the covered scene is a football game then it is probably most common to follow the action around the ball. The viewing is therefore generally guided by the structure of the event, but designing for panoramic TV demands an understanding of how this is performed.

High-resolution panoramic video

Here, we are especially interested in the use of ultrahigh-resolution panoramic video. Panoramic video is generally obtained by stitching streams from several standard high definition cameras, as in [7][10]. Importantly, stitching multiple high-definition images into a panoramic view produces an ultra-high definition image space with enough resolution to frame "virtual cameras" within the panorama, i.e. to enable interactive viewing of arbitrary region-of-interests (RoIs), produced either by the system or manually by using pan/tilt/zoom (PTZ). A single panoramic video represents the view of the scene from a given viewpoint, and the same viewpoint remains during the entire event (by panning, tilting or zooming the viewer is able to choose the viewing direction). In the case of several panoramic videos covering the same event from different points, the viewpoint might change giving users the possibility to also choose the viewing angle of the specific action.

Some systems, although providing a panoramic view of the event, are intended for a specific use, such as large screens in a cinema, as in [10], and do not offer any interaction for the users. The imLIVE [8] demo offers live streaming 360-degree video with interactivity for audiences. The panoramic camera has "only" 2400x1200 pixels, which makes it problematic to use for deep zooming. The Camargus [6], a production oriented system, offers panoramic video of the sport event created by combining an array of high-definition video feeds into one feed. The operator is given the possibility to control a virtual camera, and the recorded content is mainly used for replays.

2.2 Content interaction

When a panoramic view is combined with the tools for "looking into" it interactively, the TV viewers are given the possibility to choose their own view of the event. Emerging panoramic TV services enable navigation and interaction within the image content, in contrast to more familiar control of program features such as audio level, image controls and channel selection. Such content interaction bares a resemblance to actual production work, and accordingly it calls for extended interactive tools beyond the familiar remote control.

Flexible TV production and consumption

Recent technological advances in camera development and image processing have enabled changes in both production and viewing practices. Compared to traditional TV production where a production team produces a program – a single sequence of images selected to guide the viewer through the covered event – there is now the possibility to capture the whole event. Consequently, there are a number of systems, both academic and industrial, that enable arbitrary views of events, and flexibility for TV production and viewing. In all these systems, different types and levels of interactivity are enabled, often being very system specific. Since our concern is with panoramic television, our aim is not to cover all available systems, but to provide a brief look at relevant existing approaches focusing on key features that those systems provide. Beside panoramic television, approaches that have been developed include free viewpoint video and various forms of multi-view video.

In free viewpoint videos, researched extensively in the past decade, viewers can interactively change their viewpoint in the scene; i.e., viewers are able to freely navigate within real-world visual scenes, as known from virtual worlds in 3D computer graphics [24]. In contrast, in traditional videos, the viewpoint is chosen by the producer or director. Generally, in order to get imagery from arbitrary viewpoints, several cameras are placed around the scene, and views from real camera images near the chosen viewpoints are interpolated. Such systems enable new production as well as new possibilities for users, and are often used for sporting events, e.g. to show replays from any angle, as

in [12][13].

The LIVE project [9] attempts to give viewers their own personal and interactive broadcast by providing a production support system. The idea is to produce a parallel multistream coverage of a live event, including a backchannel for the consumers to be able to influence the broadcaster through voting [11][14]. Similarly, the My-eDirector project [10] aims to provide an interactive broadcasting service enabling end users to direct their own coverage of large athletics events and thus to take a role of a virtual director adapting the broadcast to their own viewing preferences [16][17].

User interaction with rich video content

As the above brief overview showed there are various approaches that aim to give TV viewers the possibility to choose their own view of an event, typically live sport. While there are many works describing technical approaches, there are still significantly fewer user studies that aim to show how viewers would interact with content offering various views on the event.

Olsen et al. in [15] study a prototype that offers interactive TV sport over the Internet and gives viewers both the possibility to choose the view and to control replays. Experiments showed that sport fans could easily learn the interactive controls and that they would use the interaction such as switching between cameras and ”moving in time”, rather than passively watch the broadcast. An evaluation of the My-e-Director 2012 prototype implementation of the service, offering personalization capabilities based on user profiles and recommendations, is presented in [17]. Results showed that users may become annoyed when getting continuous annotations and recommendations on switching channels.

Concerning viewers' interaction with panoramic TV, there are only a few relevant works. Neng and Chambel in [18] explore 360º hypervideos focusing on navigation and visualization mechanisms. Their “360º Hypervideo Player” provides features that assist users in orientating themselves in the panorama. The first actual user study of wide-format video is given in [19]. Authors present a study with the focus on omnidirectional video (ODV), a video that enables users to look around in 360º. The focus was on finding characteristics that make a TV-program suitable for enhancement with ODV. Interaction with panoramic video consisted of changing of the viewing angle and zooming in and out. Their findings show that ODV has the potential, yet challenges remain on a technological, content and user levels.

Bearing in mind that there are already specific services for live broadcasting where interactive panoramic video is used, there is a lack of empirical findings to inform the design of such and similar services. Today these services are not specifically requiring particular devices, which leads to a need to explore interaction with content itself. Our goal with this research is to explore how potential viewers of panoramic video can interact with the content to gain a richer viewing experience. The wide possibilities for presenting different parts of the content to viewers point to the need for thorough design guidelines in the relation to live television watching experiences. In order to take a closer look at how panoramic video content might be approached, we first conducted a preliminary focus group study. These focus groups were organized to get ideas and early input from users. Before describing the studies of interaction methods, we therefore briefly go through the method and results of the focus group study.

3. FOCUS GROUP STUDY

The goal of our focus group study was to explore potential views and preferences for interacting with the panoramic format of live television experiences. The study served as a pre-study to our further investigations of interaction methods with panoramic video content in relation to live experiences and shows. We chose a football match as the live event because the potential for following the play closely through following the ball presented a fairly simple example. Live sports were also chosen since the viewing of this type of live event requires real-time interaction, and quick decision making. In addition it was easy to elicit excitement among potential viewers and participants who were interested in football.

We organized three focus groups lasting approximately 1,5 hours each. Participants were 16 regular TV viewers, recruited from within our research organisation. The focus groups began with a general open discussion on TV viewing such as the experience of live content, followed by more concrete tasks. The material used and tasks given varied for each focus group. The participants were given a small set of tasks connected to the different materials. With the material as a helpful asset, users shared stories about their TV watching practices, which worked as a way to help these ordinary TV viewers talk about the content and presentation. All discussions were video- and/or audio recorded and the material was later analysed and structured under different key topics.

Figure 2: Group discussion around mock-ups

The first group was shown still images from a traditionally produced live broadcast of a football game. In total, 14 pictures were shown, representing various view selections like an overview of the game, an overview centered on the ball, a detail frame of a player, selected tackles, and selected audience reactions. Participants were asked to discuss the images and produce their own view by choosing the preferred views and time aligning them.

The second group was presented with keywords connected to live TV. Keywords were used to structure and stimulate the discussion regarding the concepts associated with novel TV services. We initially presented the words mobile, home, and public in order to focus on TV viewing that happens everywhere, in any environment. The words interactivity, content, multiple displays, social connectivity, and immersion represented different aspects of changing TV watching practices. Participants were instructed to choose one or two key words from each stack, and then brainstorm around the live TV topic.

The second and third group looked at mock-ups of user interfaces. The images were similar to the first focus group but presented in panoramic view in the latter sessions. Three pictures of a panoramic view in combination with additional small windows showing regions of interest and various close-ups were displayed

to the participants. Figure 2 shows a detail from one of the focus groups.

Overall, focus group participants were positive towards panoramic images of live experiences. We identified two themes through the focus groups that lead to our further development and studies of interaction approaches of the panoramic images; understanding context and social viewing.

Panorama as Understanding Context

One male participant expressed a salient need for a broader view of sports matches specifically: "*What you miss when you see sport on TV is this view (pointing on panoramic image). This is almost never shown and that is what you see when you watch it for real*". Participants expressed how the panoramic view would be able to give them a better understanding of the overall context of the live experience. They generalized across several types of content but particularly the sports events were seen as benefiting from this. One participant expressed: "*If we talk about game play, this is where we see how the game is going (pointing at panoramic image). This doesn't say anything (pointing at a close-up) about what is happening. It is like playing Monopoly and just looking at the dice*". One participant stated that he would like to have "more description" of the panorama so that he would know what he was looking at, which resonated well with others who emphasized an appropriate separation between watching detailed parts of the picture and the overview: There should be "*a balance between detail and 'where are we'*". In this case the panoramic view is then supporting an understanding for "where" we are.

Most of our focus group participants emphasized that they would appreciate to have an easily accessible panorama picture so that they could "*go back to it in one click*", an insight that went into the design of the interaction approaches that we prototyped.

Social Viewing

Another prominent aspect that came up during the focus groups was sociality. During the first focus group, the participants expressed interest in a way to coordinate with friends before the game, for example to choose the same content to watch if they were not in the same place. Doing this before the game would reduce the risk of losing focus once the live game had started. "*My feeling is that maybe it is important, this shared ... that you have the same viewpoint so that you can discuss it with your friends after watching – like 'did you see that tackle' or 'did you see that moment' – that you have shared this specific particular moment and then you have the slow motion replay of that. And you have the same kind of experience of watching*", one participant commented.

Both earlier observations at sports bars [22] and Esbjörnsson et al.'s studies [23] of sport spectators at rallies show that the gathering around a live event (watched in real life or on a screen) involves constant discussion about what is going on. In addition, our participants expressed a desire to have the possibility to share the content in other ways, for example through a web page, where they could see what others were watching and be able to switch between broadcast streams.

Based on the focus groups, we concluded that the emphasis in terms of designing for interaction with panoramic content would be to make sure dynamic viewing is supported, as well as easy access to the panoramic view.

4. INTERACTION APPROACHES

It has been suggested that rich video content calls for interaction techniques beyond remote controls. Gesture interaction and touch interaction on so called second screen devices are two broadly recognized approaches to this end. The following section presents user tests on developed prototypes for these two interaction techniques.

In the second stage of our project, two systems for interacting with such video were developed, and used in laboratory user studies to explore interactive viewing of panoramic video. The first system enables interactive panoramic TV viewing on a mobile device, either a tablet or a mobile phone, and the second one uses hand and arm gestures for interactive panoramic TV viewing in the home environment, also allowing small group interaction. These systems were developed as part of the broader project and were as such designed as probes into how viewers would interact with content produced within the framework of a panoramic, high-resolution video image.

4.1 Tablet Based Second Screen Interaction

The underlying system has been developed to record and transmit high-resolution panoramic live video in a way that allows for multiple interaction techniques on the end user side. One potential interaction method is a secondary screen that can be used to view a selected region of the panoramic image. The prototype uses an Android based application allowing for basic navigation within the panoramic picture, running on the tablet and a server [20]. Initially the client displays the full panorama. By using interactive commands the user can choose the region of interest for viewing. Interactive commands that are supported are panning, tilting and zooming (PTZ). Zooming can be controlled using pinching or using a small invisible slider bar at the left of the image. Panning/tilting is controlled by touching the screen and moving one finger around on the screen in a swiping movement.

One technical limitation is that large panoramic content (7K) cannot be shown on a mobile client, unless it is downscaled to a resolution that can be handled (e.g. 1.2K). Downscaling results in a resolution that is not high enough to see details. Thus, in the system we use here, the server crops and rescales to the resolution of the client screen (while respecting the aspect ratio of the content) and re-encodes the panoramic content based on the interactive commands (requests) from the client within a reasonable small delay, so that it is perceived as real-time.

User Study of Tablet Interaction

We conducted a study of the tablet interaction prototype in October 2012. The content used for the study was a 10 minutes panoramic video (running in the loop) of a football match that was played between Chelsea and Wolverhampton in October 2010, which was available to us (see Figure 1).

The study was conducted as a lab-based test with 16 participants between the ages of 22 to 46, with a median age of 28 years. 9 of them were not interested in football and the rest indicated an interest with one being an avid football fan. Almost half of the participants stated that they had watched panoramic video before in different contexts including large-scale movie experiences.

After a short introduction to the system, the participants were instructed to interact with the content; to try to follow the ball, look at details of their choice and enjoy the game. We recorded the interaction with two cameras, from the front and from the back. After the participant had been interacting with the system for about five minutes the researcher asked them a set of questions about their immediate experiences and impressions. Figure 4 shows a detail from the gesture study.

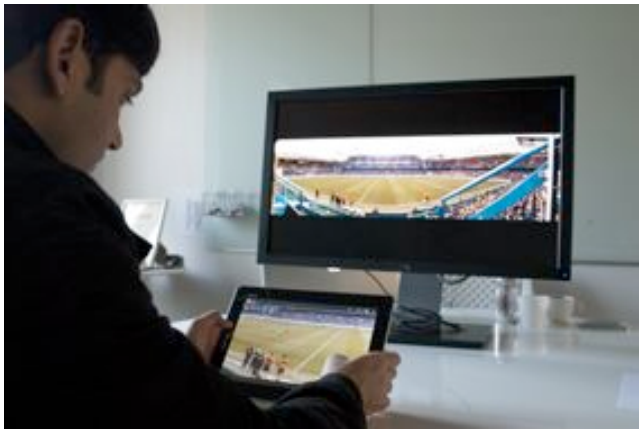

Figure 3: Interaction with panoramic video on the tablet

4.2 Gestures and Large Screen Interaction

In a second test, we tested a prototype for a large screen where interaction takes place through hand and arm gestures in front of the screen, similarly to game interaction with Microsoft's Kinect [21].

Figure 4: Gesture interaction with panoramic video

The gesture system allows the user to perform interactions with panoramic video displayed on a high definition TV screen or projected on the wall in a home environment. It works in a device-less and marker-less manner and the user is able to control the content using only their hands, either standing or sitting in a chair or sofa. The system is multi-user, i.e. a small group of users can ask for control of the system and interact with it while the others are still present in the scene.

The current implementation of the gesture system is written in C++ and runs on a single laptop needing at least 6 CPUs. The system has been split into multiple connected components where each is responsible for a single task (e.g. head location, hand tracker, gesture classification, etc.). Each component is based on SmartFlow, a piece of software developed by NIST (http://www.nist.gov/smartspace/) to facilitate the communication of software modules. The system uses the color and depth video provided by a single Kinect camera as the principal sensor. No other data provided by the Kinect is employed. The Kinect camera should be placed at heights between 120cm and 250cm and approximate at the middle of the TV screen. A free space in front of the sensor of 2-4m is also recommended.

The functionality in terms of user control of the content includes interactions such as selecting menus presented on the screen, navigating through high resolution panoramic views of the scene by panning, tilting and zooming, and control of the audio by changing the volume, muting or selecting the speaker. Figure 4 shows a detail from the gesture study.

User Study of Gestures Interaction

We conducted a lab-based study of gesture interaction with panoramic TV in a simulated home environment in November 2012. This test included two types of content (the second type of content was not available for us in the first user study). We included both the panoramic video of the aforementioned football match between but additionally, we used 5 minutes of panoramic video of a dance show where the Berlin Philharmonic Orchestra accompanied youth dancers from Sascha Waltz in the show Carmen, at the Arena Berlin in May 2012 (see Figure 5).

The study was conducted with 20 people, between the ages 20 and 39. Most of them had had experiences with panoramic videos before, either on big cinema screens or on tablets or mobile devices. 3 of them had some experience using gesture controls in this context while the other 7 were new users to this type of interaction. Where 14 of the participants interacted as pairs, six of them interacted alone. This enabled us to see any differences in dynamics when users watched and interacted with the system alone as opposed to within a social situation. All the pair users knew each other.

The participants were instructed to interact with the system, using the described gestures to zoom, pan, tilt and turn the volume up and down and mute. We video recorded their interactions from two directions, front and back, where it was possible to see both the screen and the gestures simultaneously. The participants were interacting with the system for about 10 minutes for each of two available contents, though in some cases even longer. Afterwards, the researcher asked them a set of questions about their immediate experiences and impressions, and their responses were audio recorded and transcribed.

We now continue describing the overall findings from the two studies. We keep the findings together rather than reporting individually from each study because the studies revealed comparable issues that are relevant to consider in combination.

5. FINDINGS

Both studies revealed insights into challenges and benefits of interacting with panoramic content, as well as issues regarding each method of interaction. Where the tablet study was mostly focused on detailed zooming and panning we were able to explore social interaction around live event watching in the study of gesture interaction where we observed seven couples. We first discuss four themes around our findings before presenting more specific design challenges and design implications.

5.1 Controlling the View

The main advantage to navigating (e.g. zoom and pan) within a large image is of course to be able to control the view. One of the reasons why participants want more control is that it will give them a personal view of the live show/event: "*It is always the wrong pictures they are showing, showing something completely different, and then you see something like 'that was really weird' or if you think the referee made some mistake and you want to watch it and form your own opinion,*" said one participant from the tablet study after talking approvingly about the navigation possibility on the smaller screen. Most sports viewers know where to look on the screen, which was also observed during our earlier studies in sports bars [22] and essentially supports our observation

Figure 5: Screenshot from dance panoramic video

that participants knew where they wanted to pan the zoomed-in section of the image. They acknowledged that it can be useful to control their own view, as one of the gesture study participants expressed: "*For this it makes sense if you have a lot of things going on and you want to follow*". However, there was clearly a limit to the desire to control and constantly having to actively choose the view: "*I think it's nice to be able to zoom, but I was thinking about constantly doing that. I'd love that sometimes somebody does it for me*", said another participant from the same study.

When we discussed how interactive panoramic video would change television watching, one participant said that it would give her a "*responsibility to take action*". Still, one main concern was the stress of being given too many options. "*If you are your own producer you know that you will miss something and when you have the choice of being very individual in picking some specific scene then I think it's easy for you to feel worried, like 'am I missing something? Is everyone else watching something else?'*". According to these users, having too many options can also affect the experience of the content in a negative way.

In terms of using the larger screen where the full panoramic content was running during the tablet test, participants found it useful to get an overview; one participant explains: "*If I cannot see the ball when I move the picture [on the tablet] then I look up [at the big panoramic image] to see where the ball is. Even if the ball is here and I want to see the whole image, so it quite depends. I think it is quite important to have this overview. You can see the players, where they stand. And the two teams. (...) I guess the position of the players is quite important*" This illustrates well how part of controlling the view was also about being able to view the show "from above". The panorama was thereby giving viewers the visual experience of seeing the game from the grandstand, "(…) *I like it, it is really cool to see the whole field, it is like I'm sitting there*.", said one of the tablet study participants.

Controlling the view by zooming in and out was highly appreciated. It was used by all participants in the interaction user studies and it quickly became natural for them to zoom in and out, although it was not trivial to master technically. However, some participants were worried about content being lost. "*There is a danger with all this zooming in because you lose the overview and then it just takes a few seconds and 'oh where are they now'*". The issue here is how to get the best information, and the zooming is seen as a delicate instrument for getting more details suiting personal needs. Constantly zooming in and out creates a distraction in the interaction with the panoramic content. "*It would be good to be able to move between regions of interest without having to zoom out to the panorama again"*, one participant commented, suggesting that a feature similar to a cut between cameras would be useful.

5.2 Social viewing

As we learned in our focus group study, potential viewers are keen on social functionality when it comes to interacting with panoramic content and televised live events. Although only the gestures study was conducted with pairs, it was clear that such interaction with the content enabled social interaction and could be used in a setting of two or more viewers.

However, social interaction was not always viewed in a positive light. Another aspect that emerged from our studies is that interacting too much with the content reduces possibility to communicate with your friends or family when watching happens at the same location. One of the participants in the tablet study said: "*The system is not social, when watching with friends interaction takes too much attention so you are not able to interact with friends.*" On the other hand, it can also be a reason to start discussion: "*Watching with friends (...) if panoramic content could be time shifted, rewound, on the tablet, to show friends and comment, and have a main broadcast on the TV*."

We found that the relation between what people said they prefer in terms of social viewing and how they interacted together was in some cases contradictory; where they explicitly asked for social functionality such as embedded communication, they were also demonstrating a need to simply watch the content without having to be social around it. This brings us to the next theme, the aspect of shifting viewing between active and passive.

5.3 Active-passive viewing

One of the salient characteristics of television watching is the way viewers shift between passively watching and actively interacting (through channel surfing or more complex interaction with iTV systems); similarly we observed a laid back approach in the gestures study. All participants did, at one point or another, settle on the current view and simply observed the show, if not also leaning back in the chair. They expressed in the post-viewing interviews that such lean back behaviour is part of their television habits, that it is necessary to be able to do that to enjoy the show at certain times.

When talking about potential interaction with the system, our gesture study participants often mentioned less advanced features such as simple volume up and down and channel shifting as well as being able to rewind. This was surprising in that the system afforded a high level of detail in viewing the video content and the possibility to navigate within the larger panoramic image, than what is possible in traditional television broadcast. However, many participants in fact mentioned the downsides of this detailed interaction: they were worried they would miss out important parts of the live show, particularly in a game like football. One participant provided the example that he had lost out on a goal when he had flicked from one screen (television screen) to another (tablet screen) at home the previous night.

In the study on the tablet participants generally expressed positive attitude in interacting with the panoramic video on the tablet. People seemed more willing to interact with the panorama on the tablet than on the TV set: "*I like the panorama on the tablet. I like to zoom/pan because you are used to such interaction from using mobiles in everyday life…so it is not really for a big screen.*" Another participant added: "*Having a panorama overview on the tablet was not problematic, indeed it was calling for interaction…people are used to interaction with fingers.*" Finally interestingly, it was not considered a problem watching on a handheld device: "*People already watch different stuff on mobiles (…) I actually watched handball during Olympics with my friends*". In all, users displayed great familiarity with second screen interaction, both in terms of dividing their attention and in specific interaction techniques.

5.4 Technical Expectations

Our studies also highlighted the harsh requirements that users have on technical workings of these types of systems today.

Expectations on technical quality

Even using ultra-high definition capture, there are technical limits to the resolution that can be provided in a detailed subsection of the larger image. This most critically became evident when test participants zoomed deep into details of the image. "*The resolution is the biggest killer of the application, if I zoom I want a proper resolution, otherwise it doesn't make sense. If I could get detailed view I would zoom in more*", one of them commented. Several participants reported that they sometimes zoomed in past what they perceived as acceptable image quality. "*I wanted to look at the people but then it's out of focus, to see what they are doing.*", one viewer stated, giving a concrete example of a situation where they steered their view into a very narrow section of the scene without getting more detail in the image. The sensitivity and speed of the zoom were other factors that limited the interaction. "*I was able to follow the ball, except for the long shot, when the ball would disappear. But it is also the lag of the system, if zoom was faster, it would help.*" Hence, the allowances of the zoom control – responsiveness, speed and zoom level – are important both individually and in combination, in the experience of interactive navigation in panoramic images.

Properties of the panoramic image

In the tablet study the participants pointed out specifically how a panoramic view is lacking different angles of the camera, like "*It is quite a freedom, but I'm missing different angles like in real broadcast*". Because of the camera viewpoint being fixed, it was problematic in some situations to get a good view despite the possibility to zoom, compared to a normal broadcast of a football match, where there would be several cameras at different locations in the stadium.

Participants did not typically perceive the panorama as a scene that is "always there", when they are watching more detailed regions of the image. This may be due to the relative novelty of panoramic video in the TV context, as compared to standard-framed content, and has the consequence that viewers do not take full advantage of the panoramic scene, for attaining an overview and when selecting regions of interest.

6. DISCUSSION AND IMPLICATION FOR DESIGN

Our research goal has been to understand how to look at panoramic video, i.e. how to do content interaction focusing on the specific qualities of a panorama no matter the underlying devices and interaction techniques. Here we expand on these findings picking out features of relevance for the design of panoramic TV.

6.1 Design challenges

In our studies, we have identified four categories of challenges that emerge when interacting with panoramic video content, and that need to be attended to and supported in systems design.

Balancing active and passive viewing

One of the qualities of interactive panoramic TV content is that it offers many different viewing options within the larger image of the scene, and being a spectator is to some extent a matter of finding material you are interested in. With the possibility to choose, viewers can satisfy their own interests, which is generally appreciated. Yet, there is typically a produced program or live broadcast to return to at will. As seen in our studies, users sometimes want to take it easy and sometimes want to actively control their view of the TV content; even if most of our participants expressed an interest in having their own view of the event, they were also worried that such interaction would involve too much effort and stress, and would cause them to miss important action. Interacting with and looking into high-resolution panoramic video of live broadcasts can be affected by the viewer's specific choices of:

- **Replays.** User decisions on what and when to replay,
- **Choosing region of interest by:**
  - *Zooming.* Typically choosing close-ups or regions of interest to see, for example, facial expressions or details of occurrences,
  - *Selecting pre-defined regions of interest,* provided by the system or a producer,
  - *Selecting overview.* Having the panoramic view, e.g. an overview of the whole field, and
  - *Using navigation.* Doing virtual camerawork in the high-resolution panorama picture.

From a production point of view, all these actions mean overriding the editing decisions of the producer, in the case of a live broadcast. The two modes exist in parallel and produce two alternative views of the scene. Accordingly, managing the transitions between the main broadcast and active control of the image content is a challenge. The tolerance for errors was high among participants while controlling the view themselves, but introducing interactive control at any time may also cause disruptive breaks between the passive and active modes.

Designing interaction with live broadcast content requires detailed attention to the workings of the interaction where at the same time viewers want to have the opportunity to lean back and passively watch the action. It is imperative to support both types of behaviour within interactive television systems, though the balance can be complicated to facilitate with systems like the ones suggested.

Enabling orientation

Interaction with the panoramic picture is connected to orientation. The panorama gives the users a possibility to interact within an event that is dispersed. In the same fashion as tourist signs at scenic spots can give the visitors a more informative experience, panoramic video can work as a guiding image. At the same time,

the fixed view point was seen as a limitation by users used to the fast cuts and multiple perspectives of broadcast television.

Users were explaining the use of the panorama as a mechanism for orientation and something you could "go back to" if you loose yourself in the navigation. In that sense the panorama could be understood as an interactive map that gives you a greater understanding of the scene. One user's explanation of the panorama being a menu is enforcing that there is a certain value of organization within the bigger context.

Having the panoramic picture readily available as an overview is highly appreciated, and an important feature for relaxed exploration in panoramic ultra-high definition television formats.

Individual exploration

Knowing what kind of panorama you are looking into also sets the standard for what you can understand, for example a panoramic view of a football stadium is more easily understood by frequent visitors. Sport viewers in traditional television have knowledge on what they want to see but not the equipment to interact and change their view of the whole event. Availability of panoramic video and the freedom to choose the view of the event brings out a specific dimension of TV viewing, namely, the possibility to have different selected views of the same event. This issue is in some respect already handled today with fans at an arena having separate viewpoints. This is although something new among TV viewers, where watching the same thing means the exact same images. Giving the TV viewers a possibility to interact with a panoramic view of a live event is a way of supporting individuals to create their own broadcast.

The tablet use is enabling a scenario where the individual views of the game could be even more personalized. There is no need for negotiation with other family members or friends, whereas experience becomes social and intimate in the case of the invitation by the device owner [25], or in the case of the tablet used as a second screen.

Social aspects

One conclusion that emerges from our studies is that people like being a part of a larger context, which usually involves communication and content sharing. However, our participants also expressed an interest in having their own view of the event, clearly showing that they often know what they want to see.

The social interaction that develops around watching TV can be seen from several perspectives:

- **Collocated viewing.** A group, family or friends, watching the same content at the same place (e.g. at home or at public place such as a sports bar).
- **Remote viewing.** The already well-known problem of coordinating with friends/other viewers watching the same content but located at different places.
- **Free viewing.** A newly formulated problem of watching the same program but from possibly diverse viewpoints showing different contents.

Watching TV, and in particular sports, has always been a social experience, as seen in the cases where family or friends gather to watch TV together or when discussing TV shows 'by the water cooler' (*Did you see the game yesterday?*). With advances in communication technologies, the social component has become more complex; social TV watching is possible for remote viewers too. This is already an ongoing subject of academic research and commercial development. In [27], the role of broadcasting is redefined in sense that it is not more a plain consumable TV but it serves as a source for social interaction where socializing around the content is becoming more important than the content itself. In the same work it is also emphasized that "*Social TV aims to provide multiple remote viewers with a joint watching experience*".

A taxonomy of the social aspects of television based on the presence and type of communication is proposed in [26] suggesting that there are two dimensions of the social aspects of TV. The first dimension concerns the presence of the viewers – collocated or remote; and the second dimension concerns the type of communication between viewers – synchronous or asynchronous. High-resolution panoramic video and the freedom to choose the view of the event bring out a third dimension of co-viewing. This third dimension is concerned with the possibility that viewers will have different views of the same event, thus making social interaction more problematic.

Following the sociability heuristics proposed in [28], specifically the following one: "*Support remote as well as collocated interaction*", we add that the design of interactive panoramic TV should also provide support for social conversation for viewers watching different images of the same event.

6.2 Design Implications

Here we summarize our findings and identified design challenges into design implications. The recommendations refer to how to support viewing panoramic content and to provide the live TV viewers with a better experience, and they apply to the design of high-resolution panoramic TV applications supporting content interaction.

Supporting navigating in a panoramic scene. Individual viewing and exploration should be supported in a way that makes it intuitive, yet takes advantage of production-like tools for content interaction. Careful attention should be given to designing the trade-off between responsiveness, accuracy and speed in navigation, so that viewers can follow action within the scene without getting lost in the image. Making the comparison to manual camerawork, the user-controlled view selection through *panning* and *tilting* (moving the view selection horizontally and vertically) should be fast enough to cover relevant movement in the scene, but respond in ways that do not allow jarring and disorienting virtual camera movements. As a concrete recommendation, for most scenes on a horizontal plane, e.g. a stage or a football field, horizontal movements should be more responsive than vertical ones to produce stable moving view selections. Additionally, pre-configuration through user profiles before an event would allow users to take advantage of system-defined views in a personalized manner. If the viewers can customize their viewing preferences in advance, the interaction will be less of a burden during the event. Manual interaction would then be more balanced to occasions of special interest and would draw less attention away from the content.

Enhanced point of view. Both the literature and our observations point to the importance of the spectator's point of view in watching a panorama. In mediated panoramic events, such as the ones seen in the type of interactive television investigated here, selecting a good point of view is an important consideration. If done well, it is a powerful means of conveying the experience of "being there", as a spectator at the scene of the event. This is a quality that panoramic television is well equipped to convey, compared to other emerging formats offering e.g. free form 3D navigation and other features. Designs of panoramic television should emphasize this quality as a key feature. On the consumption side, which is the main focus here, this would imply establishing the central viewpoint early on to

novice users. The affordances of the navigation tools should be explained based on that central perspective, e.g. through the metaphor of being there on the stands of a sport event, but with a set of tools for going into details of the action.

Supporting overview and detailed view selection. To fulfill their full potential for representing a scene, panoramic TV applications should include tools to provide better overviews and relevant detailed views, as well as tools for intuitively combining and shifting between the two. Access to the panoramic overview should be quick and easy. A mechanism like that would enable viewers not to get lost, and this is important since advanced content interaction also increases the risk of getting lost. Overview shots are the default image in most live broadcasts, and serve an additional purpose in panoramic television in that they provide a reference to the image content outside of regions currently framed, whether by the producer, the system or the viewer. Our study shows that viewers do not typically perceive the panorama as a scene that is "always there", when they are watching more detailed regions of the image, suggesting unfamiliarity with the format. Having the full panoramic image readily available would thus also serve the purpose of enabling orientation in the wider scene and reminding viewers of the image space they have at their disposal. Besides navigating in the image using virtual equivalents of pan and tilt movements, *zooming* is a highly appreciated and indivisible part of interactive navigation in panoramic images. As such it should be designed in a way that enables precise and quick viewing of specific regions of interests. Analogously with navigation movements, the responsiveness, speed and smoothness of the zoom need to be carefully balanced in order to meet the expectations users have on the interaction. Zoom level (or magnification) should only go as far as the resolution of the image safely allows. Practically for most current systems in the range of 2K vertically, this would mean view selections that are significantly zoomed in but not close-ups of e.g. individual people at a large distance. Our observations suggest that the ability to perform production-like operations like zooming at all is a very attractive feature. Restricting it to a smaller range is a far better option than risking exposing viewers to low resolution image content. For parts of the scene that are outside of the main action, e.g. the ceiling of an arena, possibility to zoom in should be reduced in order to stop viewers from getting lost in unrecognizable details. Furthermore, shortcuts to specific regions, which are known to be of high interest for majority of people, like the goal area, should be provided.

Support for active/passive viewing shift. Support for lean-back viewing should be enabled by providing easy access to the professionally produced broadcast, or broadcasts by other acknowledged producers, e.g. a local event producer. We suggest that, for example, local producers could take advantage of new technology and produce content for a group of viewers with similar interests. In this case an amateur producer would be equipped with a director-like functionality, making the narrative decisions and sharing the chosen view of the event with people who expressed an interest in it (multicasting). The shared content could, for example, be viewed in a public environment, such as supporter pubs, where the produced material would be oriented toward following a specific team. This would enable group of fans to get an alternative view closer to their preferences and interests with little or no interactivity. Active control of the image content should be offered to viewers, but in non-intrusive manner emphasizing voluntary transition between the passive and active modes.

Designing for co-located and remote social viewing. Social conversation for viewers watching different images of the same event should be enabled. This can include mechanisms for, e.g., following a friend or the most popular view of the event, or sharing preferences. The former can be based on friend rating and/or voting for the most popular view. The latter can include information about whether the user wants to share the chosen view and with whom. An example is adding a possibility to share interesting views; instead of asking, as is the case now, "did you see this or that situation?" users could show it, as to some extent happens now through, e.g., YouTube. To enable the proposed social connectivity, the solution should support any of the existing communication tools.

7. CONCLUSION

Due to extensive technical developments in recent years, a new type of TV content is becoming available to viewers – rich, high-definition panoramic video. In this paper we aimed to provide an understanding of how panoramic video as the interactive material can be used as TV content, whereas we are focusing on the specific panoramic picture values and experiences around it. Interaction with panoramic video assumes looking into the panoramic image space, using production-like tools for navigating (zooming, panning, tilting). Such mass capture of TV content and its delivery to viewers in the form of high-resolution panoramic video generates a vast amount of data. It is hardly possible for all the available content to be consumed at once; hence the viewers need to find their own ways of looking into the panoramic picture. In order to understand how this available mass data can be viewed, we have investigated how future users could benefit from the available technology, but also what potential problems they might face.

Based on the user studies we have identified challenges that come with this new interactive content – balancing active and passive viewing, enabling orientation in the panoramic image space, supporting both individual exploration and social conversations around it. Designs of new services based on interactive panoramic video need to resolve this challenges and emphasize panoramic qualities not present in the traditional TV content. As a start we offer the recommendations on how to support viewing panoramic content and to provide the live TV viewers with a better experience in the form of design implications.

Future studies will orient towards the design of the second iteration of the prototypes based on the findings from this work, primarily concentrating on the orientation mechanism as a support for the individual panorama exploration, but also on providing the balance between relaxation and control.

8. ACKNOWLEDGMENTS

The authors would like to thank their colleagues in the FascinatE project for their contributions to the work reported here. We especially thank our project partners UPC (*Department of Signal Theory and Communications, Universitat Politecnica de Catalunya, Barcelona, Spain*) for providing the prototype for gesture interaction and ALU (*Bell Labs - Multimedia Technologies Domain, Alcatel-Lucent Bell NV, Antwerp, Belgium*) for providing the prototype for tablet interaction. The research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 248138. We also thank all the study volunteers.

9. REFERENCES

[1] P. Cesar and K. Chorianopoulos. 2009. The Evolution of TV Systems, Content, and Users Toward Interactivity. Found. Trends Hum.-Comput. Interact. 2, 4 (April 2009), 373-95.

[2] N. H. Lansford, Jr. and L. L. Jones. Marginal Price of Lake Recreation and Aesthetics: A Hedonic Approach, 1995. J. Agr. and Applied Econ. 27 (1), July 1995:212-223.

[3] M. J. Meitner. Scenic beauty of river views in the Grand Canyon: relating perceptual judgements to locations, 2004. Landscape and Urban Planning 68 (2004), 3-13.

[4] K. Thompson and D. Bordwell. Film History: An Introduction (third edition), 2010. McGraw Hill International Edition. New York.

[5] B. Latour and E. Hermant. Paris: Invisible City, 2006.

[6] Camargus system. www.camargus.com/ (Accessed 18.11.2010.).

[7] Hego OB1 system. www.ob1.hegogroup.com/ (Accessed 17.11.2011.).

[8] imLIVE demo. www.immersivemedia.com/markets/imLIVE/index.html (Accessed 18.11.2011.).

[9] LIVE project. http://www.ist-live.org/ (Accessed 07.10.2011.).

[10] My-edirector project. http://www.myedirector2012.eu/ (Accessed 07.10.2011.).

[11] S. Grünvogel, R. Wages, T. Bürger, and J. Zaletelj. A novel system for interactive live tv. In L. Ma, M. Rauterberg, and R. Nakatsu, editors, Entertainment Computing - ICEC 2007, volume 4740 of Lecture Notes in Computer Science, pages 193–204. Springer Berlin / Heidelberg, 2007.

[12] A. Hilton, J.-Y. Guillemaut, J. Kilner, O. Grau, and G. Thomas. Free-viewpoint video for TV sport production. Image and Geometry Processing for 3-D Cinematography, volume 5 of Geometry and Computing, pages 77–106. Springer Berlin Heidelberg, 2010.

[13] N. Inamoto and H. Saito. Virtual viewpoint replay for a soccer match by view interpolation from multiple cameras. IEEE Transactions on Multimedia, 9(6):1155–1166, 2007.

[14] J. Jiang, J. Köhler, C. M. Williams, J. Zaletelj, G. Güntner, H. Horstmann, J. Ren, J. Löffler, and Y. Weng. Live: An integrated production and feedback system for intelligent and interactive tv broadcasting. IEEE transactions on broadcasting, first, 2011.

[15] D. R. Olsen, B. Partridge, and S. Lynn. Time warp sports for internet television. ACM Trans.Comput.-Hum. Interact., 17:16:1–16:37, December 2010.

[16] C. Patrikakis, A. Pnevmatikakis, P. Chippendale, M. Nunes, R. Santos Cruz, S. Poslad, W. Zhenchen, N. Papaoulakis, and P. Papageorgiou. Direct your personal coverage of large athletic events. Multimedia, IEEE, November 2010.

[17] C. Z. Patrikakis, N. Papaoulakis, C. Stefanoudaki, A. Voulodimos, and E. Sardis. Handling multiple channel video data for personalized multimedia services: A case study on soccer games viewing. In PerCom Workshops'11, pages 561–566, 2011.

[18] Luís A. R. Neng and Teresa Chambel, Get around 360° hypervideo. In Proceedings of MindTrek '10. ACM, New York, NY, USA, 119-122, 2010.

[19] Lizzy Bleumers, Wendy Van den Broeck, Bram Lievens, and Jo Pierson, Seeing the bigger picture: a user perspective on 360° TV, In Proceedings of EuroiTV '12, ACM, New York, NY, USA, 115-124, 2012.

[20] Jean-François Macq, Nico Verzijp, Maarten Aerts, Frederik Vandeputte, Erwin Six, Demo: Omnidirectional video navigation on a tablet PC using a camera-based orientation tracker", in Proceedings of the Fifth ACM/IEEE International Conference on Distributed Smart Cameras (ICDSC 2011), Ghent, Belgium, August 2011.

[21] X. Suau, J. Ruiz-Hidalgo, and J. Casas, "Real-time head and hand tracking based on 2.5D data", IEEE Transactions on Multimedia , vol. 14, no. 3, pp. 575-585 , 2012.

[22] Oskar Juhlin and Elin Önnevall, On the relation of ordinary gestures to TV screens: General lessons for the design of collaborative interactive techniques, Accepted for publication In Proceedings CHI'13.

[23] A. Engström, M. Esbjörnsson, O. Juhlin, and C. Norlin. More TV! - support for local and collaborative production and consumption of mobile tv. In Proceedings of EuroITV, pages 173–177, 2007.

[24] A. Smolic, K. Mueller, P. Merkle, C. Fehn, P. Kauff, P. Eisert, and T. Wiegand. 3d video and free viewpoint video - technologies, applications and mpeg standards. Multimedia and Expo, IEEE International Conference on, 0:2161–2164, 2006.

[25] Y. Cui, J. Chipchase, and Y. Jung. Personal TV: A Qualitative Study of Mobile TV Users. In Cesar, P., Chorianopoulos, K. and Jensen, J. F. (eds.) Interactive TV: A Shared Experience. Proceedings of 5th European Conference, EuroITV 2007. Springer, 195--204.

[26] C. Konstantinos. Content-enriched communication supporting the social uses of TV. Communications Network, 6:23–30, 2007.

[27] R. Schatz, S. Wagner, S. Egger, and N. Jordan. Mobile tv becomes social – integrating content with communications. In Proceeding of International Conference on Information Technology Interfaces, ITI 2007, pages 263–270, 2007.

[28] D. Geerts and D. De Grooff. Supporting the social uses of television: sociability heuristics for social TV. In Proceedings of the 27th international conference on Human factors in computing systems, CHI '09, pages 595–604, New York, NY, USA, 2009.