Producing Collaborative Video: Developing an Interactive User Experience for Mobile TV

Mattias Esbjornsson, Arvid Engstrom, Oskar Juhlin
Mobility Studio, Interactive Institute
P.O. Box 1197, SE-164 26 Kista, Sweden
Telephone number, incl. country code
{mattias.esbjornsson, arvid.engstrom, oskarj}@tii.se

Mark Perry
SISCM, Brunel University,
Uxbridge, Middx, UB8 3PH, UK
Telephone number, incl. country code
mark.perry@brunel.ac.uk

ABSTRACT
This paper presents a study of professional live TV production, investigating the work and interactions between distributed camera operators and a vision mixer during an ice hockey game. Using interview and video data, we discuss the vision mixer's and camera operators' individual assignments, showing the role of video as both a topic and resource in their collaboration. Our findings are applied in a design-oriented examination into the interactive user experience of TV, and inform the development of mobile collaborative tools to support amateur live video production.

Categories and Subject Descriptors
H.5.3 [**INFORMATION INTERFACES AND PRESENTATION**]: Group and Organization Interfaces

General Terms
Design, Human Factors.

Keywords
Live TV, consumer content creation, video production, mobile technologies, design.

1. INTRODUCTION
The fast moving nature of team-based sport contributes to particular forms of video-based co-ordination and image production techniques. Understanding how these processes operate has the potential to inform design of mobile collaborative production technologies for amateurs, as well as influencing research on automated video editing. Such research mirrors and complements contemporary trends in user content creation on the Internet. These concern media production and sharing, such as blogging, photo- and video-sharing. Many of these services contain a social dimension, allowing people to comment upon, and add to each other's work. The popularity of these services illustrates how consumers are changing from being passively media consumers, to take an active part in its production process, in what McLuhan and Nevitt [15] describe as becoming 'prosumers'. We argue that current and future iTV-research can contribute to this type of collaborative user content creation.

Such thinking on user-produced content is in line with Vincent and Vincent's [25] early predictions on iTV as a totally new medium. They distinguish between iTV and traditional television, where iTV will allow consumers to produce and share media material. This shift towards consumer-based media production raises new research issues, such as how existing production practices operate, as well as spurring new design challenges. A focus on media production within iTV research, also complements the recent interest on interactivity in the consumption process, as exemplified in research on making more choices available for the consumer [11], such as affecting the programs being watched [7], making TV viewing more social [14], providing games [3], mobile television [17], and so on.

In this paper, we are specifically concerned with video production as a collaborative achievement. In this, we draw on emerging research on mobile and collaborative user content creation within the areas of Human Computer Interaction (HCI) and Computer Supported Collaborative Work (CSCW). Thus, for example, Engström *et al*. [4] present a study of video jockeys and the adjunct Swarm Cam prototype, which enables club visitors to co-produce video for a live VJ-performance. Similarly, Kirk *et al.* [12] investigate what people do with video when it comes to aspects of recording, editing and sharing. In the area of user-generated content, this turn towards examining the design and use of visual display media within HCI and CSCW sits with a broader focus on the technical editing of visual materials that builds on the capabilities of recent computing developments: powerful multimedia capabilities, high bandwidth data communication, and mobile, wireless and video-enabled devices. Nevertheless, the production processes that underlie the generation of such visual display media has been left largely unexamined, in particular where this activity involves multi-participant collaboration and where real-time, 'live', multi-camera video broadcasts occur. As we will show, these live and collaborative conditions conspire to make such video production a particularly complex activity to manage and co-ordinate.

We have therefore turned to examine the design of video and TV experiences where media consumers can become active and social producers. As a starting point for such an undertaking, we provide a study on the professional production of live sports television, where we emphasize how articulation work and topic orientation play a role in co-ordination within the production process. These findings are then used in a critical design process through which we examine how such a mobile system might support the collaborative video production of a sports event.

2. LIVE SPORTS TV PRODUCTION
Professional handbooks on TV production provide a starting point for the understanding of collaboration and coordination technology in this area [16,10,26,27]. Production facilities for live sport events are typically divided into separate rooms for production control, vision control and audio control [16]. The main direction and visual production of the show is conducted in the production control room by a vision mixer, who manages the selection of video for broadcast. The setup of the room contains a

'gallery' of video monitors displaying all image sources centred around a main broadcast monitor and a preview monitor (fig. 1) and an intercom system enabling communication between this room, the camera operators and adjacent production units. Millerson argues that the production of sport on live TV involves a degree of spontaneous selection [16], in which the vision mixer 'sits in the production control room, watching an array of picture monitors, spontaneously choosing and switching between sources.' Sports productions pose particular problems for their production crew. Because of its live-ness, and depending on the rules of the game, its location, duration and other factors, team members have to coordinate their efforts so as to cover the action in a meaningful way. Events may take place simultaneously. The totality of the action may be distributed over a large area or too fast to be covered from one angle, and thus demand a combination of coverage from selected vantage points and close-up cameras fluidly following the action. Although the demands on teams differ, some general rules apply. The vision mixer must always be able to present the best angle of the action to the viewer. The camera operators must not only cover the live action but be ready to record the unexpected as it occurs [16,24]. In live production, intercom communication is normally restricted. Camera operators receive directions verbally and through a red tally light on their cameras, indicating they are "on-air".

Figure 1: Image gallery

Live sports TV mixing focus primarily on rendering an aesthetically appealing and understandable view of the action at all times. Visually, live television production follows traditional film grammar, a system of rules for how to effectively tell a story in images. A main goal in editing is providing multiple viewpoints on the covered action, and using these to produce rhythm, balance between detail and overview, and a more dynamic and compelling sequence of images [10,27]. Some key techniques are used to maintain visual continuity and are frequently used in live sports TV production. Among the most important are: First, cutting "on the action" – a cut made in the middle of a significant action disguises the edit point. If well done, the viewer will be able to follow the flow of the action across the cut without consciously realizing a cut has occurred. Second, maintaining screen direction – all cameras should cover the action from the same side of an imaginary line perpendicular to the shooting direction of one central camera. "Crossing the line" means introducing contradicting directions in the footage. Third, avoiding similar compositions – the camera angles between two following shots should diverge. Neglecting one of these conventions may result in disruptive jump cuts, or in audience losing their orientation.

Too many shots taken from similar camera angles and a similar distance from the subject is considered to be tiring for the audience. This is typically addressed by patterning of shots [10]. A scene often opens on a wide establishing shot showing the general setting and mood, followed by more closely framed medium and close-up shots of the characters. This way the viewer gets both the overview and the emotional closeness to the characters as the scene progresses. Similarly, predictable situations that reoccur throughout the production may have predefined patterns that support editing decisions and aid the vision mixer in producing meaningful footage of the game.

3. RELATED WORK

The paper is influenced by research within three separate areas. First, there are a few academic texts focusing on coordinating practices in Live TV production and the use of broadcast technology in the production of sports [13], and the material available deals specifically with the commoditisation and commercialisation of sports. Indeed, Grunneau [6] notes that many scholarly articles on televised sport focus on their textual analysis, and not on their production techniques and situated practices that underlie their creation.

Recent work by Broth [2] does however begin to address the situated practices in the actual video production process to address how they influence broadcast outputs. Here, Broth examines the co-ordinating processes of live television broadcast production teams, using Conversation Analysis to investigate mediated workplace interaction between team members. His study shows how the interaction between the vision mixer and 'script' in the control room and the camera operators in the studio is, to a considerable extent, non-verbal, and relies on all members' ability to predict each other's actions from their current performance. He also argues that the communication that takes place between the editing studio and set is asymmetrical, in that the participants in the editing room can talk between themselves, and to the camera operators through headsets, but that for operational reasons, camera operators can only communicate through their choice of framing shots and camera movement. This form of communicative action, which Broth calls 'proposal-acceptance', is repeated throughout the duration of the production, and demonstrates the importance of maintaining a shared understanding of both the desired broadcast format and of each camera operator's functional role in the production process.

Second, another related area concerns automatic video editing. A number of automated and semi-automated editing tools have been proposed, catering to a common set of identified problems. Most notably, amateur videographers lack the time and skills to produce high quality video without lengthy episodes of uninteresting and badly captured material [1,5,9,29]. The proposed tools let the user extract edited sequences from raw material, utilising various approaches, including image analysis and automating established editing principles to discard uneven camera movement and other features deemed "unsuitable" by the system [5,9,29]. Automation of live video capture has not been explored to the same extent. However, in an interesting recent attempt, Ranjan et al. [18] present a system for automatic multicamera control allowing video capture in meeting situations. Their system leverages television production principles for camerawork and uses input from a motion tracking system and a number of microphones and utilises ideal framing, movement, timing and mixing in automated multicamera productions.

Third, a recent area focuses on mobile collaborative video production. In a recent study on investigating how teenagers used personal mobile phones for video recording, analysis showed that traditional video cameras were used relatively formulaically, while mobile phones were used more spontaneously in video capture [12]. This spontaneity was also visible in the sharing of videos, which was usually done locally and immediately after recording. Users did not see the point of manipulating the clips, as these were short snippets of action. In a technical slant on such

spontaneous capture, Tazaki [23] presents a conceptual design, InstantSharecam, which emphasizes the collaborative process in video production. She envisions a group of users, each with a video camera, simultaneously shooting and co-directing coverage of an event in real time. With some similarities Engström et al [4] presents a study on how VJs produce and mix visuals live. The study informs the design of the Swarm Cam prototype, which is intended for use in club settings, where club visitors can capture video and stream it directly to the VJ, who can merge the video into the live VJ performance. This represents an illustrative example of mobile collaborative video production.

4. METHOD AND SETTING

Data collection on the live TV production process involved a number of sources and participants, and took place during three ice hockey matches in 2007 during the end of the competitive season, all at different locations in Sweden. The majority of the empirical data collected and presented in this paper has involved ethnographic observations and video-recording within an outside broadcast studio, as well as interviews, and the analysis presented below relies on this whole empirical corpus. In addition to the studio data collection, data has been collected on the work of the remote camera operators in their rink-side positions and the footage that they deliver throughout the event. We have also examined the final product of their collaboration: the broadcast match program.

In total, the study generated a substantial body of video data. Each ice hockey match lasted for approximately 2.5 hours, and as we used two cameras, the three games resulted in over 15 hours of tape recordings (including pre-and post match event). One camera was aimed at the monitors in the control room, whilst the other was aimed at the vision mixer from the side. All participants freely agreed to their participation in data collection and we have accorded them anonymity. These recordings were repeatedly viewed in team analysis sessions, and core events transcribed and categorized. Whilst video recording is increasingly used in data collection during workplace studies in HCI and CSCW, there is, as of yet, no common standard for transcribing video recordings similar to the coding schemes used in conversation analysis [cf. 8]. Consequently, we have developed a coding scheme that accounts for the material and social details of the co-ordination and work-related activities that are pertinent to the production process of televising a live sport event.

4.1 Setting

The hockey games we studied occurred in two different arenas, but with the same broadcast bus and with the same production crew. Data collection focussed mainly on the production control room, with its gallery (figure 1) where the broadcast images were selected, in the broadcast bus situated just outside of the arena. Inside the arena, two manned cameras (C1 and C2) were positioned on the rink side and high up on the grand stand (figure 2). These cameras were fixed, but free to pan and tilt, and placed on the highest vantage point up on the grandstand to provide the best overview and least obscured close up shots possible. These camera positions provided the bulk of the footage, from wide long shots to close ups. The rink side camera (C3) moved on wheels on a platform about 3x4 meters in size, slightly elevated above the ice. With its low perspective, it covered a near-180 degree view, roughly including all face-off zones (figure 2), but was obstructed by the rink in the near corners of the field.

In interviews, a camera operators described how their tasks as functionally differentiated. C1 was the main overview camera relies upon during most of the game time. C2 and C3 provide more tightly framed detail shots of the action. Their main assignment was to “follow the puck”, but their camerawork also involves patterns around frequent an predictable events. E.g. after a goal, C2 stays on the scoring player while C3 first frames the cheering audience, then the scoring team’s bench. These patterns are ‘scripted’ and commonly understood by all of the participants involved, and ensure that the vision mixer has multiple views to select from when producing the game narrative. Accordingly, C2 and C3 are used extensively in re-occurring situations such as penalties and goals. They are also a resource to the replay operator, who puts together sequences to be shown in out-of -play situations to support the narrative.

5. ANALYSIS

One of the core activities in the collaborative production of live broadcast television, and specifically for arena-based sports such as ice hockey, is ensuring that the broadcast footage of in-play action enables viewers to understand the progress of the game as it unfolds yet at the same time to convey the impression of its ‘live-ness’ to remote viewers [24]. Shots therefore need to be selected to allow viewers to appreciate these dual concerns, and the production team need to manage the narrative of the game at the same time as producing broadcast footage that is professionally produced, i.e. that broadcast footage is relatively steady, cameras are not seen to be searching for footage, and that cuts between cameras do not occur at inappropriate moments of play. Thus, the production team needs to attend to the film ‘grammar’ noted in the literature review whilst under particularly challenging conditions of action.

Note that we distinguish the production of in-play footage from the production activities involved in game pauses, i.e. out-of-play situations. We will start by going through a number of in-play situations, and end the analysis section with an out-of-play situation. As a topic, these in-play activities are very complex, mobile and unpredictable, and these problems generate challenges both for the coordination of the production team and the maintenance of meaningful broadcast footage. At the same time, the out-of-play situations provide other challenges, where the production team has to provide live footage from situations with less action. Based on our recordings from inside the production control room, we provide an analysis of the coordination mechanisms between camera operators and the vision mixer, illustrated with data from the live broadcast production.

In order to illustrate the co-ordination of the live production process, we focus mainly on one of the fundamental features of sports broadcasts: how the production team enables a seamless broadcast narrative by alternating between overview and detail shots [10] within in-play situations. More specifically, we discuss situations where the vision mixer deselects the overview camera to broadcast a close up view of the game. To understand the ways in which this is collaboratively achieved, we focus on two types of game sequences, a) in-play situations when the players are fighting for the puck, and b) out-of-play situations falling between a referee’s calls to stop and start the game. Taken together, these situations embrace most of the broadcast. Note that we have excluded pre-produced material such as commercial spots and studio interviews in longer pre-planned breaks.

5.1 Achieving in-play framing variations

An in-play detail shot denotes an occasion where a camera operator frames a close up view during the on-going game and where the vision mixer selects it for broadcast. Our data shows that such in-play shots are rare and very brief compared to the overview shots. During most of the game, the vision mixer selects C1 (see figure 1), which provides a wide view of about a third of the ice rink at any given time. Occasionally however, she selects detailed shots of in-play situations (C2 or C3), allowing the viewers to see close-up shots of the action.

There are several reasons for broadcasting in-play detail shots. In most circumstances, the only way to understand the progress of match play is to see the interactions between players over the whole rink. Individual players' activities need to be understood in the context of the interaction of many players, the details of which are only meaningful if you have the bigger picture. This bigger picture is readily provided by C1 (see figure 2), which frames a view of the play including most of the players and the puck. Yet although this overview shows the team at work, it misses out on other aspects of team-based game play. In this respect, detail shots allow an appreciation of individual skills, emotional expressions, and so on, which can only be seen with a zoom lens. As much of the skill in ice hockey lies in one-on-one play, detailed activities of individual game play need to be shown to the viewer. Overview shots provide a poor level of detail, necessitating the vision mixer and camera operators to cooperate to provide close-up footage of the action, whilst maintaining a smooth sense of transition between these different levels of focus for broadcast.

Selection of a variety of camera angles and shots is possible because the production team coordinates their work through orientation to specific task separations between camera operators. In the case of in–play footage, there is a very basic separation in the way the camera operators frame the developing events. C1 always tries to provide a broad frame, i.e. what we refer to as the overview, whereas both C2 and C3 search for close up footage. But this functional separation between the camera operators is not sufficient to simply weave into the live broadcast as it arises, and this requires additional image assembly by the vision mixer to produce a coherent and interesting broadcast footage of the game for the audience.

In the following, we first discuss how the vision mixer makes her selection between the cameras. We then examine a specific instance of the interactions between the camera operators and the vision mixer, showing how the alignment of the camerawork accommodates the requirements of the vision mixer in the live broadcast.

5.1.1 Situational concerns in camera selection

In this section, we provide data that reveals all of the occasions where the vision mixer chooses to include a detailed view from C2 or C3 (see figure 2). This data is used to reveal the narrative and interactional concerns that emerge in the live production process. A careful analysis of our empirical material reveals 28 selections of in-play detail shots during the first period of play. When those occurrences were plotted vis-à-vis their specific location on the rink (see figure 2) we identified some common patterns in the data. First, there are the selections of C3, of which all (except on one occasion, 5) occurred immediately in front of the camera operator's position (6, 8, 16, 20, 25). We refer to this pattern as rink-side. The remaining in-play detail shots were selected from C2, and can be clearly differentiated as displaying either tackles (3, 4, 11, 12, 18, 19, 22, 23, 28), or a player from the back bringing up the puck from their own zone (1, 2, 5, 7, 9, 10, 13-15, 17, 21, 24, 26, 27). In the following discussion, we examine why these situations are selected while a diversity of other situations were not. We argue that the vision mixer's selection of a close up camera is guided by narrative concerns, as well as the practical constraints of the live production process.

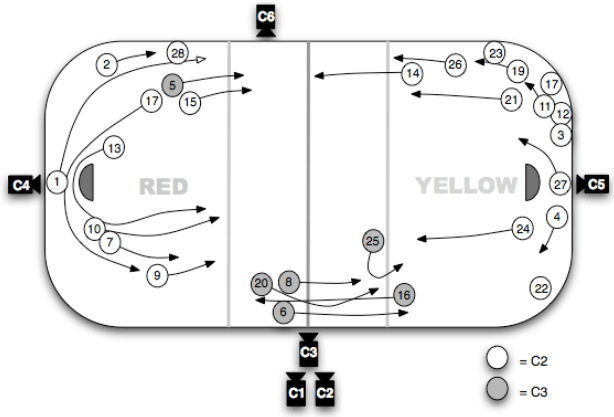


Figure 2: Selected in-play detail shots from C2 and C3

First we argue that in-play detail shots are selected when the players are moving slowly, but that this may have a potential impact on future game outcomes. If the players are moving relatively slowly, camera operators have to move the camera less than if they were skating fast. Camera operators and the vision mixer can therefore more easily predict that this particular selection can provide footage of game action that will not be lost from view.

Second, we argue that patterning, or the regularities of game behaviour, lend themselves to skilled reading of the game by the production team, and allow them to make reasonable and reliable assumptions about events in the rink, and the behaviour of the other members of their team.

Third, the empirical data suggests that to demonstrate the nature of the unfolding game it is not always appropriate for the broadcast camera to follow the puck. This can be commonly seen during tackles as well in some of the rink side selections. In those rink side selections, the puck was not visible in any of the cameras. Here, the selection of C3, showing close up footage of fast passing between players, without showing the puck, was the preferred choice. This makes explicit the demands on the production team to display the speed of the game.

Summing up, to follow the action is to broadcast footage, which includes an overview of the game, whilst at the same time, including views on individual players' actions and interactions. It is difficult to provide footage that includes both aspects of the action, and the production team constantly takes risks to broadcast such footage of the game.

5.1.2 Mediated interactional mechanisms in co-ordination

In the following section we extend this analysis and present a detailed transcript from a single tackle situation, revealing the on going collaborative achievement, i.e., how the vision mixer communicates with the camera operators, how they respond, the selection of camera for broadcast, what the different cameras

deliver, and the practical limitations on the production and selection of footage for broadcast. In the excerpt (table 1), time refers to the tape time indicator, showing minutes and seconds of footage; the column labelled "Broadcast" shows the camera currently selected for live broadcast by the vision mixer, and C1, C2 and C3 show the activities and image framing made available to the vision mixer by each of the camera operators.

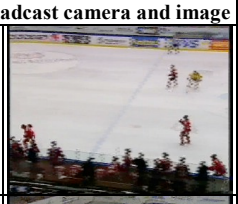
Time	Broadcast camera and image	C1	C2	C3	Verbal interaction
31:03	C1 	Overview of red team's zone on left side of the rink, showing a defender grabbing the puck.	Frames a close-up shot of the player passing the puck.	Attempts to cover passing player, but image is blocked by other players.	
31:05	C1 	The defender makes a long pass towards the yellow end of the rink. Pans right to cover the pass.	Pans right to follow the puck.	Switches framing: pans swiftly from the passing attacking player, zooms in to find the puck	
31:06	C1 	Overview of yellow team's zone: shows defending player skating to get the puck. Attacking player approaches from behind at high speed.	Switches framing: pans swiftly, zooms in on yellow player	As above	
31:08	C1 	Defender passes puck forward.	Focuses and frames on yellow defender gliding towards the rink-side. Attacking player enters the frame just before the tackle.	Zoom in and frame the defending player.	
31:09	C1 	Overview of the yellow team's zone. Attacking red player tackles the defending yellow player.	As above	Frames yellow defender.	
31:10	C2 	Moves away from the tackle to follow the puck, which leaves toward the near corner just after the impact of the tackle.	Close up on the red player tackling the yellow player Stays on the tackle situation as it dissolves.	Blocked by another player at the time of impact of the tackle. Stays on the tackle situation as it dissolves,	Vision mixer: "two, now"
31:11	C1 	Overview of the yellow team's zone (on left of camera)	Follows the attacking player while he skates away, that is not following the puck, which still is in play.	Follows the attacking player while he skates away, that is not following the puck, which still is in play.	Vision mixer: "One, now"

Table 1: Excerpt showing camera choice and interaction during in-play detail focus

This excerpt in table 1 starts when the red team is attacking. A red player passes the puck to the yellow end of the rink (31:05). The puck shoots between the defending yellow players and behind their goal cage. A red player and a yellow player then chase after the puck. The defender looks over his shoulder before decelerating at the rink side to get hold of the puck, which he then passes on. The attacking player stops skating and glides towards the defender at the rink-side, ending with giving him a hard tackle towards the rink. C1 provides overview shots throughout this sequence, in which several players and the puck are visible. C2 searches for, and provides, more detailed shots. First, he focuses on the red defence players who make the initial pass (31:03). When the puck has been passed away, he zooms out (31:06) and pans away for a four second long search, until he focuses on the defending yellow player (31:08). C2 stays with this shot throughout the tackle, but then chooses to follow the attacking player as he skates out of the situation. C3 provides the same type of shots as C2, but this shot becomes obscured by other players. The vision mixer broadcast C1, which provides an overview of the developing situation. Just as the two players collide, she says "two now" (31:10), pushes a button, and selects that camera to be broadcasted. The cut into C2 is elegantly timed in a way previously discussed as "on the action", in this case the collision between two bodies. She stays with this camera for one second to show the tackle and when the attacking player starts to skate away. She then says "one now", and selects C1 again for an

overview shot.

An important feature in the co-production is to achieve a mutual understanding between the camera operator and the vision mixer so that camera operators suspend searching for interesting topics and remain focussed on a shot for as long as it is being broadcast. In the case of studio productions, Broth [2] suggested that this kind of camera selection is handled through negotiations where camera operators propose shots by stabilising the camera on a subject. If that shot was selected, the camera operator and the director had an interactive mechanism that would support them decide how to proceed. Guided by the red 'on air' light, the camera operator would remain with that shot until the next camera was selected and their red light dimmed.

In this case, we argue that C2 proposes a shot to the vision mixer already when he focused on the defending yellow player (31:08). From that moment, C2 has the yellow player in the frame, and in focus. However, since the player is constantly moving towards the end of the rink, the camera has to move accordingly. Thus, the distinction between a moving camera and a steady camera as a means to convey the camera operator's intention to search or stay with a shot, is not possible. Still, the vision mixer identifies this as a proposition and selects it. Both players are by then at the same place for a second, until one skates away (31:11), which requires both C1 and C2 move their cameras. We argue that C2 could be selected because the vision mixer identifies a proposition given by the framing he provided, i.e. the way in which he consistently provided a stable shot on an individual player (31:08). Thus, the proposition was made because they could both recognise the content as stable and that he was going to stay with this player.

The same kind of interaction mechanism is not available when the camera operators propose and select shots during in play production, as they need to continuously follow the puck and action. Since the topic of their concern (that is the game action) is highly mobile, the camera will also be moving almost all of the time. This is the case both when it does show the action, and when the camera operator is searching for a shot. The excerpt above illustrates both these practical constraints, and how selection still is possible, without creating misunderstanding between the VM and camera operators and their subsequent broadcast of poor quality footage.

Topic oriented coordination depends on joint recognition of a specific event, in this case a tackle, and that they do this before it happens, which is empirically available in the excerpt. We argue that C2's framing of the yellow player (31:08), despite his being without the puck is done since the camera operator recognises it as being part of the game's narrative. The ways in which they "recognise" the game play as an upcoming tackle is visible since both C2 and C3 proposed a detailed shot of the defence player even before the attacking player reached him. They also both leave enough space in the frame for the attacking player to enter it before the tackle. This indicates an understanding of the event that is to come. Furthermore, the vision mixer cuts from C1 into C2 just as the bodies collide. This would be hard to do if she was not orienting towards it as an upcoming situation. It is also noteworthy that she makes the cut exactly "on the action" i.e. not before or after the tackle but just as it starts (indeed, to the precise frame).

These activities, taken together, support the argument that the organisation of the production is based on orientation towards specific topics of narrative concern. This is made possible as the camera operators and the vision mixer do not simply capture what is happening on the ice, but what is going to happen. Furthermore, their coordination based on identifying relevant topics is not only important in establishing an agreement on when camera operators should stay focused, and when they should go looking for other topics. It also includes agreeing that the temporal unfolding of the game will likely continue to yield narrative relevant material in the current framing of the game, if that camera is selected. Thus, they need to predict the temporal unfolding of the game. This follows Sudnow's [22] point about the nature of temporality in social interaction, in that it allows co-participants to interpret and orient themselves towards the 'internal time structure' of activities.

In addition, this coordination does not only include interpreting the players' intentions, but also involves managing the practical constraints of the situation at hand. When the vision mixer selected C2, the camera operator had just stopped moving and zooming. We suggest that this is of importance for the selection, but for different reasons than Broth's interpretation of the meaning of camera movements [2]. All of the participants had displayed their orientation to the game play as a topic. However, since ice-hockey is highly mobile, it might be practically difficult to provide detailed shots of action. If a player makes a pass or a shot it is very difficult for C2 and C3 to provide detailed shots of the puck being passed, and so they do not attempt to do so. We argue that C2 could be selected because it provided a rather immobile situation. Thus, the steady camera position provided by C2 was taken as a practical opportunity by the vision mixer that it would be possible to broadcast a detailed shot from.

Summing up, we argue that mixing overview shots with detail shots was made possible through mutual orientations to topics, which was possible because the temporal unfolding of actions was a recognizable feature, and because the narrative feature was only used during brief moments with a relatively immobile character. Thus it was possible to display one to one actions with more emotional features such as face expressions and body impacts. Finally, these details did not occur at the same location on the rink as the puck, showing how the complexity of the rule that camera operators should "follow the puck" is in its interpretation and use.

5.2 Managing in and out-of-play transitions

The use of detailed shots is more extended in out-of-play situations during breaks in the game called by the referee. We have analysed the vision mixer's selections during all of the occasions of a regularly used transition between out-of-play and in-play production; the face-offs taking place in the first period.

The face-off allows play to be resumed after a pause in the game. It is an important and reoccurring event in the TV production since it marks every transition between out-of-play and in-play time, and is recognisable by all members of the production team. The referee drops the puck in between two competing players in the middle of a circle on the rink. The other players wait outside the circle until the referee lets go of the puck. The regularity and fomal structure of this event lends itself to a detailed analysis of camera transitions and the cooperative mechanisms that underlie its broadcast. First, we discuss what camera is selected by the vision mixer to narrate the activity. Second, we will further examine the principles for selection by also categorising whether the shots are occluded or not.

There are 19 instances of face-offs during 20 game minutes of in-play activities, i.e. nearly one per minute of in-play time. The

face-off is divided into three temporal phases; *before*, *during* and *after* the drop. The length of these phases is between two and six seconds, depending on the unfolding of game play. We summarise these transitions in the table below:

Selection sequence	A	B	C	D	E	F
Before	C1	C2	C2	C1	C3	C2
During	C1	C2	C1	C3	C2	C2
After	C1	C1	C1	C1	C1	C2
Number of occurrences	6	2	4	3	3	1

Table 2. Vision mixer's selection of cameras in face offs

Table 2 displays the ways in which the vision mixer selects cameras during face-offs. Alternative A, where the mixer stays with the overview camera (C1), is the most common. But in all, staying with one single camera (sequence A and F) is done only in seven out of 19 occasions. Alterations between different cameras (B, C, D and E) is done in the remaining twelve occasions. Thus, switching between overview and detail is the preferred choice in this type of transition to in-play situations.

These selections can be broken down further, and in table 3 we discuss those selection sequences where the vision mixer alternates between cameras. We argue that her decision is influenced both by attending to the game's dynamics, but also by practical necessities of providing an unobscured view of the puck and the players in its vicinity. We further categorize the footage provided by the cameras as being either directly relevant to display the situation or not. 'Irrelevant' shots provide no imagery of the unfolding face-off, but may still be relevant to the game as a whole.

	Camera	Occasions selected	Occasions selected whence providing irrelevant content	Occasions delivering relevant material being non-selected	Total number of occasions providing relevant content
Before:	C1	9 (47%)	-	10 (53%)	19 (100%)
	C2	7 (37%)	2	4 (21%)	9 (47%)
	C3	3 (16%)	3	5 (26%)	5 (26%)
During:	C1	10 (53%)	-	9 (47%)	19 (100%)
	C2	6 (32%)	-	9 (47%)	15 (79%)
	C3	3 (16%)	-	6 (32%)	9 (47%)
After:	C1	18 (95%)	-	1 (5%)	19 (100%)
	C2	1 (5%)	-	7 (37%)	8 (42%)
	C3	- -	-	4 (21%)	4 (21%)

Table 3: the relevance of live broadcast footage

In the table above, we summarise the vision mixer's selection of broadcast camera, as the face-off unfolds. Additionally, the table reveals if the different camera operators deliver relevant live footage during the three phases of the face-off, which the vision mixer can select for the live broadcast.

In all cases, C1 delivers relevant footage, and is chosen in half of the situations occurring before and during the actual face-offs. Accordingly, C2 and C3, assigned to provide close-ups, are chosen in general every other time, much more frequently than during in-play time. Why is it that the vision mixer suddenly gets this preference for blending detail with overview? As with in-play situations, we argue that it is influenced by the mobility of the topic at hand. A face-off is a relatively stable situation, where the players are positioned for the drop, waiting for the puck hits the ice. Being a fixed and reoccurring feature of the game, it is largely predictable and arguably less interesting than in-play action. Thus, although C1 is a safe choice and will always provide relevant footage , camera variation is preferred.

After the face-off, the selections of broadcast camera are completely different: C1 is selected in all face-offs except one.[1] C2 delivers relevant content in only seven occasions, and it is selected at only one of these. C3, in turn, is never selected immediately after the face-off, and only delivers relevant material in 4 occasions (out of 19). This illustrates how the vision mixer's selection is sometimes made out of pure necessity, in situations where the overview camera is the only safe alternative to fall back on.

The occurrence of occluded shots might be explained by the specific characteristics of the situation immediately after the face-off. This is highly mobile, and neither the camera operators nor the mixer can be certain in which direction the puck and the players will go. The situation is the most difficult for C3, which is positioned lower than the other cameras, which might explain why it provides less relevant footage although its viewpoint gives it certain qualities when it comes to deliver close-ups of the action.

Summing up, out-of-play situations and transitions differ from in-play situations in their lower pace. In those situations, the vision mixer makes more use of balancing overview with detailed shots, always relying on camera one as a safe fall back. Still, the game is complex and mobile even in these occasions.. The cameras who are supposed to deliver a detailed view also often find themselves with occluded views after they have framed a specific topic. Again, this make the gesture interaction that Broth [2] suggests inappropriate as there is no possibility of holding the camera still. Instead it points to the importance of topic orientation in the production.

6. DISCUSSION

The study is not intended to influence professional TV-production. Instead we aim to influence research on automated video editing and mobile collaborative editing of video streams. We consider two areas that we offer a contribution to below:

Automatic Editing systems: Co-ordination of live multi-camera systems and video mixing process is highly demanding, even for professionals with years of experience. Computer technology offers opportunities to automatically edit live video streams

[1] On the one occasion that C1 is not selected, the vision mixer chose C2. This face-off is an exception from all others in the first period because the referee blew the whistle just before he drops the puck. He replaces the player from the yellow team taking part in the face-off, and C2 follows this player with a close-up when he slowly skates away.

together, much in the way that desktop video editing systems such as Adobe Premiere and Apple iMovie do for offline consumer video editing. Recently, we have seen this design approach extended to incorporate live video editing from several video streams to provide variations in between detail shots and overview shots [18]. Although, this approach appears promising in the constrained setting of an office meeting, our analysis would suggest that such design approaches will struggle with complex settings and camera configurations such as those seen at sports events. For example, we have discussed how the production depends on recognition of emerging topics by the production team. Such 'topic oriented' co-ordination during the game means that the production team has to recognize the action on the ice as being of a particular type of activity, and then to orient towards this activity in guiding their own subsequent activities if they are to present a narrative structure that will help the viewers appreciate the overall picture of play or the performance of individual players. This is necessarily a skilled practice, and requires a deep understanding of the game. Furthermore, these topical situations are generally extremely brief and need to be recognized as likely to take place before they actually occur in order to select an appropriate and timely in-detail shot. Thus, the team has to orient towards emerging topics, rather than just what is currently happening on the ice. Furthermore, the vision mixer has to be aware of the possibility that the shot, when framed by the camera men, become occluded and therefore become much less relevant for broad cast. This makes automation incredibly hard to achieve, and as 'intelligent' image analysis technology is still in its infancy, it will be unlikely to offer much support in these circumstances.

The analysis also demonstrates the interpretative flexibility around the rules that the production team follows – or rather, as we emphasize, the rules that they orient towards. In this, there is an inherently situated dimension to creating a narrative through the choice of broadcast video during the game. For example, in the interviews we were told the general rule for filming was to follow the puck. However, the activity pattern that we see during the tackle clearly diverges from this. Here, the cameras stay on the tackling player through the tackle, and only then move on to display the puck. This also happens occasionally in the 'rink-side' pattern. Clearly, there are other concerns in editing than just showing the puck that interest the production team. Thus, there are rules that are somewhat conflicting, and there is an interpretative flexibility of what to show in a given situation. This also makes automated editing problematic, in that rule-based process are largely inflexible, as it is not necessarily appropriate simply to follow the puck: an automated editing system that tracked and broadcast video footage showing the puck would be inappropriate in many such instances.

Yet particular areas and design directions in the live editing process could be augmented or supported through partial automation. One example of this arises from our analysis, which reveals how the vision mixer's selection of in play detail shots are patterned to an extent, which suggests a foundation on which to implement some sort of automatic editing technology. Although they are not absolute rules that are invariably followed, regularities in the editing process, such as those observed in the in-detail framing and selection of slow moving players, could be brought to the attention of the video editor as a suggestion that they choose these shots when the broadcast had settled on an overview for a long period of time. Taken together, the intricate nature of a sports event, with its complex social interaction, offers very practical problem that make it difficult to produce through automated means, although there are clearly opportunities here to offer live video editors support in making broadcast decisions.

Collaborative Mobile Editing As well as providing support for automated editing, it is also possible to envisage how networked mobile devices might support the collaborative editing of concurrent video streams by non-professional camera users. These might include situations such as the broadcast of live images of motor sports by ringside fans, or of soccer matches, by parents. However, it is far from a trivial design challenge since live production is highly time-constrained; with rapidly moving and dynamic game to take to air, asymmetrical patterns of collaborative interaction and resources for communication between the remote camera operators and vision mixer, and a demand to produce televisually exciting material that we have seen in the fieldwork. Indeed, we have shown how the mixing of overview shots with detail shots in play requires enormously skilful action. The vision mixer does not select in-detail shots whenever possible, but only under particular conditions in which they are narratively relevant, and practically possible to co-ordinate without camera jumps. The use of mobile collaborative video editing begs the question how this collaborative orientation to specific topics, such as a tackle, is supported. How editors ensure that the amateur camera users they are working with will follow the same events is a very real problem in these settings – after all, there may be many topics of interest within a sports event, from video of their family and the setting itself in addition to the sport. Nevertheless, here too, there are opportunities for design.

As our analysis shows, mixing between overview shots and in-detail shots depends on a clear functional separation between camera operators in framing the topic at hand. In professional production this is relatively straightforward in that their tasks are pre-allocated, but for amateurs, this may be more flexible and open to negotiation. Teams might to decide on tasks before the event, in line with the professional team. However, and depending on how formal or stable these video collaborations are, participants might need to do ad hoc negotiations on this during the production, and might require dynamic allocations as users entered and left the location or moved around within it. Whilst a basic and important feature in the coordination of TV production is the stable and known positions of the cameras, in an amateur production the positions and availabilities of camera may have to be negotiated and articulated. Here, we can see value in supporting the articulation work around how requests for a particular type of footage might be dynamically sought out or allocated.

The data also illustrates how much of the interaction between the vision mixer and the camera operator is oriented to acquiring agreements on what the operator intends or should do next, i.e. staying with the current footage or to search out a new topic. We have shown how the complex and mobile character of the topic at hand provides for a very lean and brief form of interaction, akin to 'grabbing' a camera briefly for an in-detail broadcast. We suggest that this form of interaction might also be applicable to collaborating amateurs. Thus a person making a live broadcast could 'grab' remote camera footage providing a detailed shot for a brief moment, whilst indicating to the camera operator that this was in progress with a signal such as a tally light to indicate that they were currently 'live'. This would not require camera users to

negotiate or articulate the shot proposals or selections through more heavyweight co-ordination mechanisms in these instances.

7. DESIGN BRIEF

In the following, we outline a mobile system supporting collaborative video production, oriented towards sport events. We have investigated how professional TV-production is performed and we now turn to how this production process could be achieved through mobile collaborative video technologies.

We take the SwarmCam system [4] as a starting point. It provides live streaming of video from Symbian S60 mobile phones to a web interface, drawing on an open source program called Movino [21]. Quicktime components, on a VJ mixer program (based on max/msp/jitter) running on a laptop, receive the incoming video streams. The mixer consists of a user interface displaying preview windows of the incoming streams and controls for a basic set of mixing functionalities and effects such as brightness, contrast and hue controls, and tools for spatial montage. The interface also contains an output window, equivalent to a program monitor, which can be set to full screen mode or output to a separate screen. In the following we discuss alterations required to provide mobile and collaborative amateur sports video production.

Most pertinently, the editor/vision mixer interface needs to be more simple and specific for the topic at hand. We suggest that it should only contain four windows displaying incoming data, as well as quick keys associated to each of these to provide standard cut transitions. Our basic concept is that all components of the production process should be achievable on a mobile phone, hence the camera operators should record with their camera phones, and the vision mixer should be able to chose a live stream on the mobile phone. Accordingly, the SwarmCam system needs some critical additions, as discussed below:

Support for articulation work As argued, the system must support for more extended social interaction than professional systems, seeing that an ad-hoc group of amateur collaborators cannot rely on pre-defined roles, tasks, or camera positions. Accordingly, the SwarmCam system needs to be complemented with various forms of support for communication between the mixer and the camera operators. First, we need to consider adding an audio channel where they can converse, in a way that is parallel to the audio track of the topic at hand. However, this audio channel for articulation might depend on the same bandwidth as the capturing of video and sounds of the game. Bandwidth limitations are probably the most critical constrain for this type of mobile applications. Therefore, we may need to design other ways of supporting this social interaction.

Second, we suggest that SwarmCam is complemented with a more elaborated tally light. Instead of using only a red light when the camera is on air, we could support communication with extended graphic capabilities and text. The text communication could be either pre-set messages such as on-air details, overview, speed, or text chat support. We suggest that pre-edited text messages are suitable for a camera operator, whereas a mixer might also be able to write short messages to camera operators. As we discussed previously, the social interaction might concern much more than in professional production. It might range from discussing topic coordination, to combining video production with other relevant tasks. The latter can include socializing with other members in the audience or just taking brakes. It is essential that the system supports even these types of coordination work.

Finally, we noted that the location of the camera operators, which is pre-configured in professional TV production, will be much more negotiable among amateurs. They might place themselves in awkward positions from a narrative perspective. They might also have to reposition themselves to accomplish variations in framing, similar to how the rally audience select their positions to acquire overview or detailed views. For technical reasons, detailed framing might only be possible on a mobile phone by being positioned close to the topic. Furthermore, repositioning might be necessary when renegotiating topics. In all, an amateur mixer will be less certain of the geographical positioning of her camera operators. Therefore, we suggest that this automatically communicated through the system. For outdoor events, this could be done with GPS-technologies that are increasingly available on mobile devices. This would allow the vision mixer the possibility of making an initial selection based on the positions of each camera operator.

Support for topic orientation As discussed earlier in the analysis, the ability to orient to topical elements in the game is an essential skill that professional camera operators develop. The camera operators frame events that are loosely centred around the main game play, but also include more peripheral topics in order to assist the vision mixer in her assembly of a larger game narrative. Footage of the audience, referee, coaches and players in the penalty booth is thus weaved into the mix to support the production of the broadcast of the game. This relies mainly on the camera operators' common understanding and ability to predict the game, and to a much lesser extent on verbal instructions from the vision mixer during the live production. In fact, restrictions on verbal communication during the fast unfolding of the game make it necessary for the team to base their work on pre-defined patterns and topics.

In designing collaborative support for mobile video production this problem persists, and at least some basic skills of professional camera operators need to be transferred to non-professional users. In our design, a degree of predetermined *functional separation* is implicit in the system. In order to take part of a SwarmCam mobile production each participant assumes a given role: Overview Camera (OC), Detail Camera (DC) or Vision Mixer (VM). These are simplified roles corresponding to the professional tasks described in the analysis, each with a basic set of instructions to make the collaborative production possible to conduct. These instructions include the basic tasks of each role during game play. There is thus a necessity to incorporate some form of negotiation of these roles within the system.

Pattern templates, expanding on the basic instructions for the tasks of each role, could also add professional qualities to the collaborative production effort. For instance, the ice hockey camera operators' patterns in given situations could be transferred to similar production situations. In this way, the system could be design to accommodate increasing levels of commitment by gradually adding pattern templates to the basic methodology as the production teams' ambition increases.

8. CONCLUSION

The design of user experiences for future TV lies as much in the hands of amateurs as in the hands of the professionals. We suggest that the recent boom in individual content creation on the web will find new and innovative forms. We have in this paper suggested that future technical support for production might be more collectively organized. Based on a detailed ethnographic study of

professional TV production we identify both possibilities and challenges in such a perspective.

The highly time-limited conditions of sports TV production, with fast moving and unpredictable players, combined with the demands of a live broadcast, sets specific challenges on how the situated coordination between the remote collaborators, i.e. vision mixer and camera operators, is pursued. In our findings we emphasize how the live video stream is used both as a topic and resource for collaboration: whilst it forms the nature of the work, it is also the primary resource for supporting mutual orientation and negotiating shot transitions between remote participants. The vision mixer communicates with the camera operators through short utterances over radio. These utterances are heard by all camera operators, additionally they can see if they are chosen through the red tally light on their camera, and they can switch on a display between their live footage and what is currently broadcasted. They communicate back to the vision mixer by attending to recognisable "topics" of live footage. It is also necessary to predict them before they occur, to adopt the production to the fast moving game. Hence, coordination is performed with a minimum of rich interactive sequences and few utterances.

These findings have distinct implications which could spur the design of mobile collaborative video production tools. In particular we suggest a focus on tools to support the articulation of the organization, both in pre-production and during the production. Such tools, which are already somewhat available in a TV crew, will be more essential for the amateurs than for the experienced professional.

9. REFERENCES

[1] Adams, B., Venkatesh, S. (2005). IMCE: Integrated Media Creation Environment. Transactions on Multimedia Computing, Communications and Applications. ACM, 1 (3), pp. 211-247.

[2] Broth, M. (2004). The Production of a live TV–interview through mediated interaction. Proceedings of. Int. Conference on Logic and Methodology. SISWO..

[3] Chorianopoulos, K. and Lekakos, G. (2007). Learn and play with interactive TV. *Comput. Entertain.* 5, 2 (Apr. 2007), 4.

[4] Engström, A., Esbjörnsson, M. and Juhlin, O. (2008). Mobile Collaborative Live Video Mixing. To appear in Proceedings of MobileHCI 2008.

[5] Girgensohn, A., Boreczky, J., Chiu, P. (2000). A semi-automatic Approach to Home Video Editing. Proceedings of UIST. ACM, pp. 81-89.

[6] Gruneau, R. (1989). Making spectacle: A case study in television sports production. In L. A. Wenner (Ed.), Media, Sports and Society (pp. 134-154). Sage.

[7] Hand, S. and Varan, D. (2007). Exploring the Effects of Interactivity in Television Drama. Proceedings of EuroITV 2007. Springer-Verlag, pp.57-65.

[8] Hindmarsh, J. Heath C. vom Lehn, D. and Cleverly, J. (2002). Creating Assemblies: Aboard the Ghost Ship. Proceedings of CSCW'02. ACM., pp.. 156-165.

[9] Hua, X.-S., Lu, L. and Zhang, H-J. (2003). AVE—Automated Home Video Editing. In Proceedings of Multimedia'03. ACM Press. pp.. 490–497.

[10] Holland, P. (2000). The Television Handbook, 2nd ed. London: Routledge.

[11] Jensen, J. F. (2005). Interactive television: new genres, new format, new content. In *Proceedings of the Australasian Conference on interactive Entertainment* Creativity & Cognition Studios Press, pp.. 89-96.

[12] Kirk, D., Sellen, A., Harper, R., and Wood, K. (2007). Understanding videowork. Proceedings of CHI'07. ACM, pp.. 61-70.

[13] Krein, M.A. and Martin, S. (2006). 60 Seconds to Air: Television Sports Production Basics and Research Review, In Arthur A Raney, Jennings Bryant (eds) Handbook of Sports And Media. LEA: NJ.

[14] Luyten, K., Thys, K., Huypens, S., and Coninx, K. (2006). Telebuddies: social stitching with interactive television. In CHI '06 Extended Abstracts on Human Factors in Computing Systems. ACM, pp.. 1049-1054.

[15] McLuhan, M. and Nevitt, B. (1972). Take today; the executive as dropout. Harcourt Brace Jovanovich.

[16] Millerson, G. (1999). Television Production, 13th ed. Woburn, MA: Focal Press.

[17] Oksman, V., Noppari, E. Tammela, A., Mäkinen, M. and Ollikainen, V. (2007). Mobile TV in Everyday Life Contexts – Individual Entertainment or Shared Experiences? In Proceedings of EuroITV 2007. Springer-link, pp. 215-225.

[18] Ranjan, A., Birnholtz, J. and Balakrishnan, R. (2008). Improving meeting capture by applying television production principles with audio and motion detection. Proceedings of CHI'08. ACM., pp. 227-236.

[19] Real, M.R. (1975) Super Bowl: Mythic Spectacle, Journal of Communication 25 (1), 31–43. 20.

[20] Silk, M., Slack, T., and Amis, J. (2000) Bread, butter, and gravy: an institutional approach to televised sport production. Culture, Sport, Society, 3, 1-21.

[21] Storsjö, M. (2007). The design of Movino - S60 phone client, OS X components and video server. http://www.movin.o.org/ (Accessed 3rd, Feb, 2008).

[22] Sudnow, D. (1972). Temporal Parameters of Interpersonal Observation. In Ed. David Sudnow Studies of Social Interaction. New York Free Press.

[23] Tazaki, A. (2006).. InstantShareCam: Turning Users From Passive Media Consumers to Active Media Producers. Presented at the Workshop Investigating new user experience challenges in iTV: mobility & sociability, at CHI'06.

[24] Verna, T. (1987). Live TV: an inside look at directing and producing. Focal Press: London.

[25] Vincent, G. and Vincent, F. (1996). The Future for Interactive Television. iTV'96 conference at the University of Edinburgh.

[26] Ward, P. (2000a). TV technical operations. Woburn, MA: Focal Press.

[27] Ward, P. (2000b). Digital video camerawork. Woburn, MA: Focal Press.

[28] Williams, B.R. (1977). The structure of televised football, Journal of Communication 27 (3), 133-139.

[29] Yip, S., Leu, E. and Howe, H. (2003). The Automatic Video Editor. Proceedings of Multimedia. ACM. pp. 596-597.