Amateur Vision and Recreational Orientation: Creating Live Video Together

Arvid Engström[1] **Mark Perry**[2,1] **Oskar Juhlin**[1]
[1] MobileLife@Interactive Institute, Sweden
{oskarj, arvid.engstrom}@tii.se
[2] pandi, DISC, Brunel University, UK
mark.perry@brunel.ac.uk

ABSTRACT
We explore the use of a live video broadcast system by a group of amateur camera operators to film an event on networked cameraphones. Using an interaction analysis of physical interactions and orientations to the work of others, we examine their choice of camera angles and positions in their filming as they attempt to provide interesting visual content and a coherent narrative. Findings illustrate how users adapt their behaviour as co-ordination problems occur by drawing from a set of everyday visual practices ('amateur vision'). They also show how the specifically temporal aspect of live video requires extended attention on its production, and that this is at odds with the 'recreational orientation' of amateur film crews who simultaneously participate in events for their own enjoyment and film them on behalf of other viewers. Implications for the design of collaborative live broadcast media are made, focusing on approaches to interaction design that augment users' visual practices and allow users to look on behalf of others while experiencing places and events themselves.

Author Keywords
Collaboration, user-generated content, broadcast, live video, embodied interaction, leisure, amateur, professional

ACM Classification Keywords
H5.2. Information interfaces and presentation.

General Terms: Human Factors, Theory

INTRODUCTION
Recent years have seen a number of parallel trends in how video content is produced, distributed and consumed on the Internet. Enabled by high speed networks and advances in consumer video technology and smartphones, video now makes up over half of the traffic online, and is forecast to rise to 91% by 2014 [5]. A key component of this traffic flows from sites like YouTube, Vimeo and Facebook, largely powered by user-generated content. This content is evolving from single-user, single-clip uploads to *collaboratively produced* rich media (e.g Qwiki, Kaltura). It is also is increasingly offering the opportunity for *live* broadcast (Qik, Livestream, YouTube Live), and this has been applied in areas as diverse as live community and crowdsourced news broadcasts, sports TV and emergency management [6,8,14]. Alongside these developments, an emerging field of research addresses the interactional and experiential challenges of these new video formats [1,17,22,24]. As well as their impact on video consumption, these emerging technologies and platforms offer research challenges for the production of video media.

This paper adds to the research on video interaction by exploring issues in the production of video through these new technologies by presenting a field trial of a prototype system–the Instant Broadcasting System–that embodies a combination of *collaboration* and *liveness*. The IBS allows its operators to create user-generated content, namely broadcastable live video footage from multiple networked cameraphones. It represents a new genre of tools for collaborative video production among non-professional users. The issue of concern for us is that professional video production involves both technological and practical skills that are not easily transferred to amateurs, in addition to filming taking place in a very different organisational context. To explore this, we draw on a naturalistic field trial and describe some practical challenges that occur for users of the IBS. Focusing on the ways that amateur camera operators do 'looking together' to generate complementary video streams, their problems in applying their everyday visual practices to understanding camerawork ('amateur vision'), and their participatory immersion in events ('recreational orientation'), we then discuss implications for the design of technology for collaborative production and advanced user generated content.

One of the benefits of collaborative video production is that content can be captured on multiple devices, through their multiple perspectives on a topic, and these video streams can be edited live into a more visually interesting story. This obviates returning to the material at a later time to perform a secondary task of editing it *after* the event [16,24]. Live editing also has the advantage of immediacy, allowing content to be fed into broadcasts as events unfold, as well as giving multiple visual perspectives on an event to remote viewers. For professional broadcasters this is a key value, but it remains to be seen how such value can be realised for *amateur* productions. Much of the emphasis on non-professionally produced visual media in commercial applications and research has been on single camera productions and where the action is not live (such as YouTube). However, although collaboration could potentially support more advanced forms of broadcasting, collaborative work is rarely unproblematic, and our analysis indicates that multicamera video production is no different.

CSCW 2012, February 11–15, 2012, Seattle, Washington, USA.

Looking at future directions for these trends in production practices online, the addition of advanced forms of collaboration has great potential for new forms of content [23]. However, alongside this we can expect familiar CSCW challenges, as previously professional practices and technology are transferred to wider user groups. We have already seen game-changing uses of *mass online collaboration* (most notably Wikipedia), but have yet to see successful examples of *live collaboration* as described in this paper. The live element of broadcast video, with no opportunity to re-edit content, adds another layer of complication to its production. At some point, as collaboration becomes more demanding, contributing to the best of one's ability (e.g as Wikipedia author) may not be enough to participate in a meaningful way in the new practices envisioned. Time-critical collaboration in live video may be close to that point. When designing future live media for group action, we need to understand these issues better, and our study offers insights into this topic.

Video in everyday life has received some attention in the research literature, most notably in Kirk *et al.* [16], which explores how video is produced, edited and consumed with existing devices. It makes evident how video production is an explicitly *social* process in all of these aspects of its use [see also 1,18,24], showing how its *end use* is a key driver in video production, be this to share 'in the moment' (typically on 'lightweight' devices, such as cameraphones) or for more meaningful events seen later on (on 'heavyweight' video recorders and used to create DVDs). In adding *collaboration* and *liveness*, live multicamera productions go beyond these everyday video practices; they are themselves social activities, with the added complexity of time-critical collaboration on a common visual product. But live video productions also support and enhance important qualities observed in Kirk's *et al.* work; they allow sharing in the moment and remove the need for cumbersome editing after the event.

The context of the user study presented in this paper is a music festival, and the video content created by the camera operators was broadcast onto public screens in front of a large audience between music sets. The role of the camera operators was to capture different and complimentary visual perspectives on the event on behalf of a vision mixer (the person assembling these multiple viewpoints into a narrative by selecting feeds to broadcast). Our study shows that the capture of multiple shot framings and angles is problematised by the ways that non-professional camera operators understand the production process, collaborate together, and rely on an ordinary understanding of visual media that is not well suited to filming.

While the familiar limitations of mobile networks and small format devices do feature to some extent as constraints on behaviour in the user study, detailed analysis shows that a focus on amateur collaborations that involve multi-camera filming brings new issues to the fore. These issues are different to those observed in the practices and problems faced by professional operators in live multi-camera broadcasts [cf. 7,21]. There is more to the difference between on the one hand video, and on the other text and photography, than the issue of technical literacy and skills. Live video production is an inherently time-based media, which requires a more extended attention to the production during the event than these other media. This requirement makes explicit a consideration of the different motivations between amateurs as media producers and professional media producers, what we term their "recreational orientation." We frame the issues observed with the amateur camera operators as a set of shared and embodied visual practices, and show how these have implications for the design of real-time video-based user-generated content.

MEDIATED LOOKING BY AMATEURS

Video technologies support a type of collaborative gaze in which camera users act as proxy viewers on behalf of someone else: the eventual viewer of broadcast content. You could call this type of camerawork 'mediated looking'. It involves looking editorially (making decisions on what to select for display), looking together as a team (in providing one of several complimentary camera perspectives), and looking on behalf of others (primarily the viewers, but also others in the production process, such as the person mixing the camera feeds). Obvious problems that might be faced by camera operators when they perform this mediated looking are, a) they may not look at what viewers are interested in, and b) where several camera operators are involved, they may not organise themselves in a way that provides an angle or position of camera gaze that complements the video streams provided by other camera operators. The latter is important for the person choosing which images to broadcast. These are problems faced by professional as well as amateur camera operators, but professional operators work within a community of practice that draws from a common understanding of the video production process, as well as organisational protocols, technical skills and knowledge that the amateur operators cannot exploit in their filming. Collaborative camerawork in a setting with limited means for communication relies on a shared set of expectations about how others in the team will act and make sense what of they are doing. Whereas professional operators have a rich set of knowledge and practices, amateurs must draw from other organizational resources. To complicate this, amateur camera operators may have other interests and concerns that impede their actions and interactions. We therefore turn to examine technologically mediated looking, and how non-professional users experience and make use of visual technologies.

The focus on *seeing* in research on collaborative systems has previously been on the ways in which people piece together visual observation and other resources to accomplish work. Seeing (and more broadly, awareness) is considered an important feature of work: it is important to look around to understand what is going on, where, and who is doing what. For example, the importance of

supporting awareness, through making shared objects, such as flight strips [e.g. 20] visibly observable for team members has become a cornerstone of CSCW research. In this regard, there has been a focus on embodied interaction when sharing perspectives between collaborating participants, because of the ways that people are able to orient their bodily positions, gaze and manipulate objects, not just in order to do their individual tasks, but to make the nature of their ongoing work available to others.

Taking the point that amateur collaborative camerawork is not simply looking together, but working together to produce visual content, we turn to discuss this topic. How amateurs understand the way that work is visually made available to others is a topic of concern when non-professionals work together to create content. However, the terms 'professional' and 'amateur' are problematic, and views on amateur production tend to be skewed towards the notion of users' lack of expertise and knowledge. The distinction between them often focuses on skills and being a competent practitioner [cf. 15]. Thus professionals possess exclusive domain knowledge, while amateurs are seen either as inept, or having romantic qualities of authenticity. However, the wide dissemination of digital production technology has blurred these categories and brought a range of content that does not easily fit one or the other [4, 23]. Aspects other than skill may be as important, most notably for this paper, the amateur might have entirely different motives for doing a particular task than a professional. For professionals, performing the task is a source of income, whereas the amateur might be doing it primarily for a sense of belonging to a social group, or for their own enjoyment. For these reasons, we have used the term 'amateur' (in inverted commas), for lack of a better term, to denote practitioners with diverse skill sets and motivations, partly or entirely outside of traditional 'professional' categories.

If recreational enjoyment is a key motivation, then previous studies of enjoyment practices can be seen as resources for understanding this topic. For e.g., tourism often involves finding out what to do and where to do it, often in collaboration with others. In this regard, Brown and Chalmers [3] have shown how tourists balance the requirement of solving that task through a combination of techniques including the use of tools such as maps and guidebooks, but also through paying attention to making it as enjoyable as possible. The focus on enjoyment was, for example, visible in the way users perform scheduling in an ad hoc manner, and in the way that they prefer working on vague or crude resources, rather than detailed information; so for e.g. they might head out for areas with plenty of cafés rather than navigating to a specific shop, because of the likelihood of interesting encounters, rather than explicitly planning an efficient, goal-oriented excursion, but with the same overall intentions of going shopping.

Juhlin and Weilenmann [13] also show how users balance efficiency and enjoyment, but in a very different setting. Their studies of collaborative, non-professional deer hunting reveal how the recreational hunters' requirements for enjoying their day on the hunt and getting to their prey was organised into a rhythmic variation between concentration and relief, and between solitude and socialisation. Being an amateur, it appears, is not just about being inexpert, but is also managing a balancing act between enjoyment and action. To be an amateur camera operator is likely to be no different to these other domains, and of particular interest to UI designers, user actions performed on devices and the ways that they deploy devices may be as much to do with their engagement with the setting as it is with the task of camerawork.

THE INSTANT BROADCASTING SYSTEM PROTOTYPE

The Instant Broadcasting System (IBS) examined in this user study is composed of a *mobile client application* and a *server side application*. The channel between the two applications is provided by a mobile network, and in this study, operated over a private WLAN. The mobile client, Movino, is an open source application that captures a video stream from the built-in camera on a mobile telephone (Nokia N85, Symbian series 60), encodes it and streams it over TCP/IP [movino.org]. Originally conceived as a single-user application, we have further modified Movino to support communication between camera operators and a remote vision mixer. In terms of its interface, the Movino application shows camera feed, with the addition of a number indicating data connectivity, a red light indicating the camera is selected, and simple onscreen text messages.

Figure 1: Vision mixer interface to the IBS (part)

The server side application is a program that consists of an interface displaying windows of the incoming video streams (up to 5, see figure 1, bottom), and controls for mixing and other functionalities (not shown). In fig 1, three live cameras are ordered sequentially, with camera 3 currently selected to broadcast 'live' onto the Internet. The interface contains a 'live' window (top left, equivalent to a professional 'program' window) and 'preview' window (top right). Just visible at the bottom of fig 1, the interface also provided backchannel interaction from the mixer to the camera operators. This backchannel allowed the vision mixer (VM) operating it to send text messages to the

camera operators, and to 'buzz' them with the vibration ringer on the cameraphones.

METHOD AND SETTING

Using empirical data of a team of amateur camera operators, we examined how they collaboratively created visual content for live broadcast using the Instant Broadcasting System. Data collection took place in Sweden during summer 2009 at a music festival. Two groups of camera operators were studied, one Swedish and one Dutch, aged 17-19. Both groups had previously undertaken a short introductory training course in video, photography and storytelling media production, so understood videocamera operation and live video mixing, although this was not at a sophisticated level. Both teams made a number of broadcasts using three cameras, typically of two standard formats: interview and action coverage. Camera operators were themselves filmed and screen captures were made of the vision mixer's monitors. We debriefed the participants immediately after filming, with a 30-minute group interview with camera operators and the vision mixer (VM); these were recorded and transcribed.

The field trial gave us a rich understanding of how the participants planned and performed the broadcasts. In the analysis that follows, we present material from both the debriefing discussions following filming, and on one filming episode by the Dutch team. The use of single data point, as in the latter case, is typical of interaction analysis [e.g. 10]. Whilst it may not cover the full breadth of behaviours, it allows us to examine–in depth–the impact of practices, technologies and contextual features that impact on social interaction and media production. We recognise that it can be hard to generalise from a single case study, but this nevertheless stands as a perspicuous study [9]–one that focuses attention on features around important aspects of the collaborative production of amateur live video.

Before being issued with cameras, we gave the operators a 10 minute background introduction to the project. They met the VM, and were given basic instructions on using the cameraphones. They also received basic instructions on filming, i.e. to try to provide steady shots and provide footage that would complement that of the other operators, in camera angles and framing. Although relatively little training or instruction was given, this was not a technically complex or sophisticated system to operate. But as we shall show, using the system in collaborative filming proved to be a practical challenge for other reasons.

In the tables of data that follow, the first column shows the time from the start of video data collection. The second shows any visibly identifiable actions by the camera operators or significant relevant environmental events. The third column shows the current broadcast camera (identified as C1, 2 or 3), and if they occur, whether the camera has 'hung' and fails to broadcast while showing a still image displayed onscreen, or becomes active again; this is visible on our screen capture from the VM's computer and the public screens that are visible on the researcher's video recordings. Notably the camera operators can also tell that a camera feed has been dropped or hung by attending to the public displays or via their device's data connection icon.

In the textual explanations of the data (second column), 'forwards' and 'backwards' refer (respectively) to movement towards the front and rear of the basketball court. Simultaneously timed events are coupled within time boxes below to aid interpretation: they may, or may not, be connected in other ways. Arrows in the diagrams indicate paths and directions of movement. Finally, the 'ray' lines from the camera operators in the diagrams show the approximate perspectives as seen from the participants' cameras. Diagrams (figs 3 to 6), rather than photos, have been used because the camera operators and their visual orientations are extremely hard to ascertain from a static screenshot, and because these images are visually very dense, making annotations on the images hard to discern.

ANALYSIS

This case illustrates a situation where collaboration failed and as such, provides a critical point for drawing implications for design. It starts off by the camera operators providing a variety of shots to the VM, but these soon change to very similar camera angles. This sequence lasted around 5½ minutes. Prior to this broadcast, the Dutch team had filmed a ten-minute interview. For this, their second broadcast, they were asked to film an action sequence of their choice, without interviewing and focusing on camera angles and imagery. They selected a temporary basketball arena, around 20 metres directly in front of the stage and public screens (fig. 2).

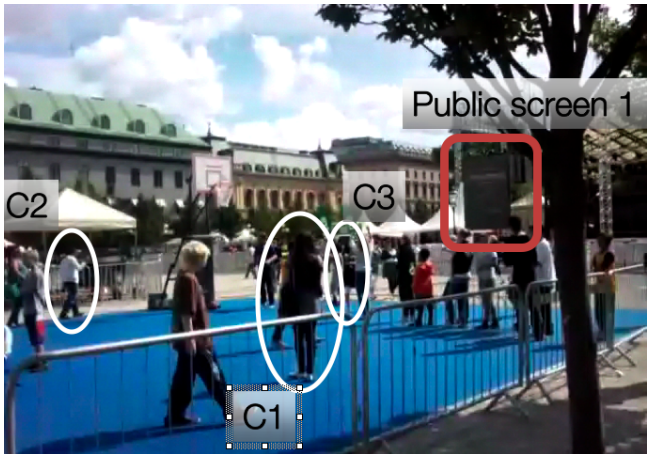

Figure 2. Arena, starting positions of the camera operators (numbered) and public screen

The backchannel malfunctioned during this particular sequence, which hampered their communication with the VM pre-broadcast. This might seem to have been a serious issue for co-ordination, as the design of support of communication between the VM and the camera operator has been suggested as a critical aspect in designing multicamera systems for amateurs [21]. However, the text tool proved problematic in filming. During the debriefing sessions, the operators commented extensively on this, with

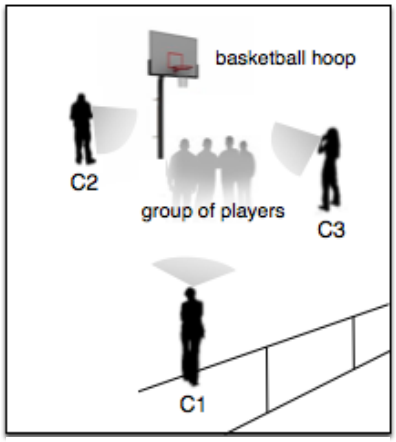

Figure 3. Setting, user positions and orientations prior to broadcast

comments such as *"In the beginning I got them, but when I was interviewing I didn't have really the time to check if she was messaging something"* being typical. Messaging also proved to be something of a distraction*: "You get a little more focused on what will come and a little less on what you're filming*. There was a broad consensus that the text backchannel was good for co-ordination up to the point of going live, but less useful while broadcasting. The iconic red *tally light*, meaning a camera is on air, was effective. Not relying on the text backchannel, the teams' organisation and camerawork was based on the on-going events around themselves, their orientation to the work of the other operators and to the live broadcast on the public screens. To support their collaboration with one another whilst filming, the camera operators were able to orient towards both the physical positions of the others (by attending to their spatial locations, bodily orientation and gaze) and what was currently being broadcast (through two large public screens either side of the stage area showing the live broadcast).

The participants had been in position for about 10 minutes, waiting for their broadcast to go live on the screens by the stage (see figs 2 and 3 for their relative positions and camera orientations). The transcriptions and discussion around the data is split into two phases to help structure the dense data. Although the observed action is continuous and there are no formal boundaries between stages, phases can be loosely considered to be composed as being topically distinct. In practice, the phases distinguish substantial spatial shifts of the participants around the arena, and segment the data into more easily referenced elements. In phase one, the camera operators remain largely in the same locations that they initially selected for their shots. Phase two begins with a rapid and extended movement by one of the camera operators, and involves the other camera operators attending to this reconfiguration in their formation. In the end all the three operators move around, adjusting their positions following intrusive movements from the others.

Phase one – Steady positions

Our analysis begins as the camera team of three took positions at the basketball court, awaiting their broadcast being shown on the public displays adjacent to the court (see figure 3). After having waited in place for over ten minutes, the live broadcast began with an inauspicious start: the selected camera (C3) showed a frozen image of the basketball game on the public displays. Indeed, screenshots from the VM show that all of the camera feeds were frozen. 11 seconds later, the cameras unfroze, and the mixer went on to alternate between two of the cameras. Over the next few minutes, the mixer broadcast all three cameras, but also had to manage the camera feeds occasionally freezing up. The camera operators largely kept to their original positions, only adjusting their viewpoints slightly, and attending to the resulting large public broadcast displays.

What we can observe in the data here is the VM initially working with poor broadcast footage, before settling into a sequence of paired shots from C2 and C3: long shots from behind the basket and close up shots of the players making the shots. During this period, it is visible from the data that C3 appears to be actively using the public screen as a feedback resource to check whether she is being broadcast. Indeed, in the debriefing interview all participants report that they do this to see who has been selected for broadcast.

The configuration of the camera operators in this phase provided multiple, complementary views of the game. Operators were dispersed over the site and not coming into shot or physically blocking each other's footage. We interpret this as that they were following the instructions they were given at the start of the event to shoot different angles and provide a mixture of close-up and overview footage. During this phase they step back and forwards, oscillating around a small area (see fig. 4, a). Their behaviour is similar to that of professional operators who take fixed camera positions and act according to operational conventions that guide them to frame particular topics and shot types [21]. We suggest that these micro-adjustments occur because they are searching for an acceptable shot within the instructions given. However, phase I ends with a marked change to this, led by the actions of C2.

Phase two – A Camera operator leaves his position

In this phase, we see a sudden disruption in the organisation of the camera operators, caused by C2 leaving his position. He circles the arena and eventually blocks another camera operator. The other camera operators both become aware of this disruption, and try to adapt to their unique spatial positions being usurped and their views of the action being blocked. The analysis and C2's own comments below provide alternative explanations for his experimentation with the organisational format of filming.

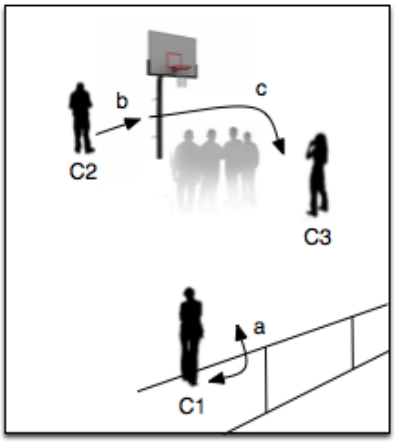

Figure 4: user movements in phase II

So, this phase starts

with C1 still selected, when C2 takes off, moving fast along **b** (fig 4), behind the basketball basket and towards C3's position. He pauses briefly, then moves further towards C3, circling behind, then around the other side of her. During all this time, C2 is not selected:

13.04	C2 begins to walk fast along **b** (see fig. 4)	
13.11	C2 circles behind basketball basket (see fig 4, along **c**), and stops	
13.12	C1 shifts position back along **a**	
13.14-13.28	C2 walks further down towards **c** (fig. 4) beside C3, then walks behind her (**d**, fig. 5)	C3 (@13.20)
13.25-13.40	C1 steps right 2 steps, then moves back against fence	
13.30	C3 glances quickly to her right	C3 hangs
13.33	C3 glances to her left (in direction of C2)	C1
13.37	C3 turns her head up and to her right at the basket	
13.38	C3 glances very quickly to right, then left towards direction of big screen	
13.40	C3 glances at C2 to her left again	C3

So above, we see C2 move around the basketball court to take up a position beside C3 to produce an almost identical perspective to her, and this is noticed: twice, C3 makes clearly visible glances towards C2 in his new position (13.30 and 13.40). The action sequence continues as C3 moves to avoid being hit by a basketball, but then continues to move away from her original position, passing C2 (ending at 13.44). C2 then moves again, stopping directly in front of C1. During all this time, C2 is not selected for another 16 seconds, and is only selected when the footage from the other cameras is unavailable for technical reasons. C1 then moves forward (**f**) and as she pauses, C3 begins her move to the rear of the court (**g**):

13.43-14.00	C3 jumps left to avoid being hit by ball, and continues to move to the back of the group of players and towards the rear of the court	C3 hangs (@13.45) C1@13.46 C1 hangs (@13.465)
13.44	C3 steps left and towards back of basketball court, filming the basket to her right	
13.48	C3 steps backwards	
13.49-13.53	C2 moves forward fast (see fig. 5, along **e**), stopping to block C1's view at 13.53 (C2 is several metres in front of C1, facing away from her)	C3@13.49 (v. poor image)
14.06		C2
14.12-14.22	C1 steps forward towards (**f**), stops at 14.22	C3@14.19
14.22	C3 edges towards the rear of the court (**g**)	C3 hangs

Perhaps the reason for the movement by C1 and C3 here is that during this time, cameras 1 and 3 lose their network connection and hang, and there is no reason for them to maintain a steady shots for broadcast, allowing them the freedom to move around before they come back online. Nevertheless the movement from C2 during phase II is odd, especially given the duplication of his perspective with C3 (starting at the end of phase I) and his physical blocking of C1's footage (13.53). The final part of the sequence is excluded for brevity. Then all three operators move around as they make adjustments to their positions following the changes imposed on them by the others, and they end up blocking two of the three shots provided to the VM.

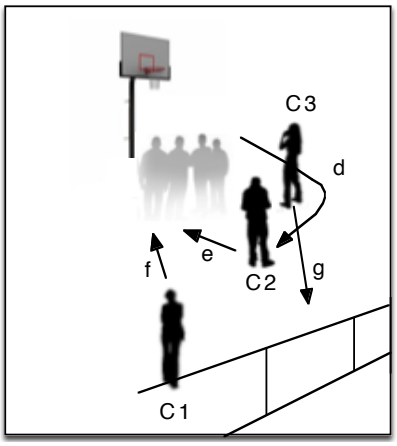

Figure 5: user movements in phase II

The most analytically interesting aspects of this phase consist of C2 moving off and C3's reaction to this duplication of her perspective and subsequently, C2's intrusion into her field of view. We can see from the data that C3 glances at C2 several times (13.33 and 13.40). She also glances towards the screen. Thus, she has a 'peripheral awareness' of what is going on (i.e. of C2's actions and the footage being broadcast). We therefore interpret her move (14.22) as a reaction to C2 moving into her area and thus providing a similar shot to her one. This interpretation is confirmed in the debriefing interview:

C3: I really wanted to zoom in on people. I started with that and then I looked around me and they were also close to the people and then I backed away and sat on the ground [...]
Interviewer: You were very close
C3: Yeah. I wanted to stay there but it was boring because they also came in very close. So I had to go away

Here, C3 wanted to provide a close up (or 'zoom in' as she refers to walking up towards people), but when she saw the other crew members, she moved away. Thus, her movement was motivated by the need to provide complementary and unique shots. The question we must ask here is what motivated C2 to impinge on the others' footage? Was it that he thought he was not providing the VM with interesting material? Looking at the actual broadcasts, although he had not been selected for 58 seconds when he began his move at 13.04, he had been selected three times already. That might tell us that he had another motivation than being receptive to the VM. Whatever the reason, we see C2 moving near to C3 to give the same shot angle, despite their initial instructions to provide complementary footage. The reasons for this are partially explained in the debriefing interview:

C2: Because nobody was making close-ups. Everybody went to the back
Interviewer: You said that nobody was making close-ups?
C3: I was standing in their faces!
C1: Ha ha she was
C2: and then you went back

Here, C2's first explanation is that he was trying to make a complementary shot to C3. C2 justifies this by inaccurately claiming that C3 was not making an appropriate close-up shot. It could be that he wanted to follow the filming

instructions and act according to their initial setup, but then he did not think that C3 was delivering appropriate footage and tried to make up for her inadequate camera work, i.e. that he did not trust her to provide such material. Second, when challenged on this by C3 and C1 he explains why he started to move, by orienting to another principle of organising filming, that is, through an imperative for mobility in camerawork as the way to make a "picture", perhaps referring to the ways that professional camera operators use tracking shots:

C2: I think everybody must walk around. Then in order to [pause] moves like a picture [pause] if you not move. You must move. But not too much."

Although this reference to 'a picture' seems unclear (he is not a native English speaker), he goes on to describe that this references TV and film imagery. C2's explanation for leaving his initial position is both to back up another (in his opinion) failing camera operator, and to provide a new type of footage that is complimentary to footage from the other cameras. C2's explanation may be a *post hoc* rationalisation of his actions, but his justification provides us with insights into how he understands the role of camerawork, and when certain patterns of action become appropriate. We cannot discount the possibility that C2 gradually begins to follow other criteria in his filming that were not agreed upon beforehand or shared with the others and just starts to film as if he were on his own. Indeed, this is a strong possibility. In the following section, evidence emerges that C2 is not attending to the activities or positions of the other camera operators as he claimed.

What then is happening? There appears to be a contradiction between C2's organizational principles and individualistic behaviour, and this can be understood as a shift occurring over time. This is seen in that C2 shifts to another principle during filming, clearly not in understanding the others. This indicates that he is discarding his own footage over time and selecting a more preferable role–that of a moving camera. Thus, the sequence shows that he acts on individual preferences not shared by the others. So, it would appear that there are a number of organizational principles in camerawork that the participants reflexively orient towards. This does not mean that they always follow them, although where those rules are shared, the participants need to demonstrate that they are being attended to. This is exactly the point that we see in C2's explanations of his behaviour–he recognised that there were organisational principles that he had an obligation to orient to (in his interview explanations), yet this is very evidently not observed in the empirical data. This tension between teamwork and an individualistic perspective (i.e. as 'event vision') is a serious problem for camera teams and for future technology designers.

DISCUSSION AND IMPLICATIONS FOR DESIGN

The task of collaboratively producing a compound visual story out of a number of individual camera feeds is essentially the same whether it is done by a professional crew or a group of 'amateurs'; the topic of the story should be covered in a diversity of angles and sizes of framing, and made available to the VM who assembles the broadcast in real time. In this sense, the multicamera format dictates a practice that is essentially different from everyday forms of 'snapshot' video use observed in earlier work [16,18]. It would perhaps be beside the point to hold 'amateur' cameraphone teams, whose motivation may not be to produce high-end broadcasts, up to professional standards. But keeping professional organisational as a point of reference in how 'amateurs' go about managing this task, highlights some distinctive characteristics of collaborative production. Here, we discuss three of these: the roles of *sensemaking and organisational adaptation* around the perceived actions of others, the role of individual *visual practices* in social organisation, and *experiential tensions* arising during collaboration. We will show that these characteristics cannot simply be written off as lack of participant thought or effort, and discuss their implications for the design of future technology for collaborative production of live media for non-professional users.

1. Loose collaboration: shifting positions and roles

While professionals organise their camerawork in preset ways and rely on skilled workmanship to achieve this, 'amateurs' will usually be more loosely organised. In the broadcast situation examined here, we see how the team of three are able to quite comfortably emulate a basic camera setup after some brief instructions. But as time passes, it becomes clear that they do not have a clear or common idea of how to continue providing good footage and what organisational schemes to follow in their collaboration. Each individual's camerawork needs to dynamically adapt to what the other camera operators are doing, so that they each produce different angles and framing as their common topic changes over time. This is visible in the ways in which they check on the other operators and move in search for variation. This peripheral awareness–alternating between looking at the camera interface and a gaze taking in the other camera team members' actions–is an important resource in their collaboration. However, they do not only adapt the way they perform their roles. They also seem to shift role as part of this adaptation. C2 appears to simultaneously work with two organisational principles, in 1) backing up for another failing camera operator (as he describes it) and 2) that all of the operators should move constantly. It is clear from the video material and interviews that they were following several organisational principles. This coordination issue led them from initially filming multiple visual perspectives to providing similar viewpoints. Notably, there was no practical way to verbally communicate these principles during the ongoing filming.

A communication backchannel would appear to be a partial solution to this problem, but presents several challenges. An audio backchannel, as seen in professional systems, has the drawback of being unidirectional, as camera operators cannot reply to the VM verbally while filming. Indeed, text

messages, as implemented in the IBS, quickly proved to be of limited use. Although not operational in the particular situation analysed here, the debriefing interviews also showed that while text messages were useful for pre-planning, they were too attentionally demanding for both senders and recipients for effective real-time interaction and organisational adaptation during broadcasts. Adding the ability for camera operators to speak over the backchannel would therefore seem inappropriate, and indeed might become cacophonous as the number of cameras scales up.

Design points: One of the obvious problems faced by the camera operators is that they have to be aware of intrusions into their own space, so that their shots are not physically blocked or that they share the same camera framing or perspective as the other users. For example, local proximity sensors (e.g. bluetooth) coupled with directional indicators (such an integrated compass) would allow the system to make its users aware of the presence of their collaborators encroaching into their space and filming in the same direction whist they were engaged in looking through the viewfinder. This could be similarly achieved with simple image processing on the server side. However, heavyweight technical approaches to co-ordinate camera positioning would seem unnecessary, as it is not easily possible to ascertain that similar perspectives are indeed shot duplications, moreover, the data shows that camera operators were at lest partially able to manage their own recovery from spatial intrusions by adjusting their positions and shot framings. It may be that simply making participants aware of such breakdowns would offer the possibility of socially organized resolution as the participants became aware of developing problems. Similarly, our video and interview data has shown that individual users' awareness of their own, and others', contribution to the production as a whole is important for effective collaboration. Support for such peripheral awareness could include a switch to view the broadcast program at any time. The camera's red 'tally light' proved to be instantly understandable and also served as a motivation to seek interesting content to film (although at the expense of adding a competitive edge to get 'the best' shot). This positive feedback could be extended to display total use of one's camera 'on air' over time and relative to other team members.

2. Ways of seeing: professional vs. amateur vision

While the work of the production team is to produce a coherent montage (i.e. create different footage that is cut together to form a continuous sequence) and the camera operators strive to do this, their lack of organisational skills and *individual* 'ways of seeing' appears to hinder their collaborative ability to practically achieve this. This is a special case of special case of organisational difficulty, and develops one of the issues around loose collaboration further: we call this *amateur vision*, because, in contrast to Goodwin's *professional vision*, the camera operators have no corresponding shared "socially organized perceptual framework(s)" [11: 616]. Each camera operator carries their own subjective interpretation of what categories of shot is interesting to view and what to highlight as relevant and meaningful when choosing what to film in combining their own footage with that of the others in their team [cf. 7]. We have deliberately not overtheorised 'amateur vision' as a conceptual contribution but a matter of practical concern for the participants. Our point is that the participants do not have the skills of seeing in a professional way for making video [cf. 21], and rely on skills drawn from their experience of other everyday activities. Thus, on his 'seeing' of the developing situation, C2's vernacular skills provide an imperative for mobility that is evidently inappropriate, in that it blocks the other camera operators' views, provides an unstable platform for filming and removes the opportunity to select shots from his original position. This can be seen in his interview data, when he describes how you must 'move like a picture'. It is not that team members are inexpert (they have many relevant and partially relevant skills), but that they visually analyse and orient to the setting as a resource for collaborative action using everyday sensemaking skills.

This is in clear contrast with the work of professional camera operators; the role of professional vision in video camerawork has been discussed by Macbeth, in what is described as 'the praxeology of seeing with a camera' ([19:152]. Given that the participants in this study have no encultured common practices or shared ways of seeing the visual landscape as a team, they must rely on their own existing visual practices, most likely drawn from, on the one hand, watching visual media, such as television, and on the other, producing visual media, such as photography or making home videos. In producing the footage, the production team (i.e. cameras and VM) are working to provide a sense of scenic intelligibility [12] in the reverse way to that of a viewer, who interprets the film "*much as we would understand the order and properties of the everyday social and natural world… Our understanding of a film text trades off our knowledge of the structures of everyday activity and practical reasoning"* [8:289].

Borrowing from Garfinkel [9:227], these observations act as an aid to our sluggish imagination, 'reflections through which the strangeness of an obstinately familiar world can be detected', and our findings uncover what now seems an obvious, yet important point. 'Amateur' crews *can have no* skilled or commonly-known-to-be-shared understandings of how to film beyond those they have discussed already, and they can have had no way to suddenly develop these in situ without further communication. We argue that the camera crew are drawing from a combination of vernacular visio-cultural practices (as competent members of a visually literate society) and their prior experience in viewing, filming and photography. Further, given their different experiences and the complexity of their task and setting, these are highly likely to be understood and enacted in a unique way by each individual. Hence, C2 surprised the other team members by apparently shifting to a different role during the broadcast. This role was not apparently arbitrary or constituting a lack of skill. But it was subjectively applied, uncalled for, and in

conflict with the other team members' filming. Such individual practices may not facilitate the collaborative production of multicamera video, and herein lies a problem for design: how to develop methods of interaction that will allow camera teams to collaborate effectively in the absence of a shared, socially organized perceptual framework.

Design point: Non-professionals cannot rely on their training and encultured common understandings to change their practice as the unexpected happens. They may well be orienting to completely different notions about what to do and how the other participants will make sense of their own actions. Some micro-coordination can be done locally, but it is reasonable to expect that they should rely more on direction from the VM. To this end, this is a case where an audio backchannel could provide a practical addition for time-critical coordination by the VM. For a basic setup, or where the addition of headsets would be impractical, text communication could serve as a fallback. Such verbal interaction is also more sociable, thus offering a motivation for 'amateur' user involvement, and builds on existing conventions in co-ordinating online gaming.

3. Enjoyment: recreational orientation

The 'amateur' teams observed were doing this filming both for their own enjoyment and are also engaged in an enjoyable setting. So, whilst the collaborative effort of filming offers a reason for focusing on their task of creating video for remote broadcast, they were also oriented to, and engaged in, pleasurable aspects of the very event that they were filming. This leisure, or recreational orientation provides a conflicting set of attentional demands on them. Paradoxically, this personal enjoyment is the very reason for them being there so that they are able to do the filming. While it may seem to be a weakness for effective collaborative production and high production values, this recreational-orientation also offers valuable opportunities for collaborative video systems. First, when enjoying an event, users may be more motivated to record aspects of it for posterity or to show it to others. Second, they are likely to know what is interesting to broadcast–after all, this is why they are likely to be present at the event. Third, interesting leisure activities are likely to appeal to enough people to provide the likelihood of multiple camera users attending to offer different and complementary views of the event. This mix of motivation, topical expertise and critical mass are powerful tools to harness in collaborative video.

Furthermore, the role of the professional camera operator, as we have previously argued, is to look at something on behalf of someone else, i.e. the VM and the TV-audience. We argue that the 'amateurs' have another motivation in that they are looking on behalf of others while experiencing the event for themselves; combining the camera vision they adopt in their role when filming with an event vision. This dual motivation might explain C2's actions in moving around to both film similar angles and block the views of the other operators. That is, camera operators also look at the topic that they are filming as someone attending an event: the *camera vision* becomes confounded with the *event vision*, which implies strolling around and looking at events as an ordinary participant at the event or a tourist. Thus C2's concern for mobility may be as much a concern to please himself as to please those he is filming on behalf of. Moving around is a way to get a better view of the event for himself. So for the professionals, while looking with a camera is not quite the same as just looking [19], this may be a less relevant concern to amateur operators. Indeed, the way the participants describe the footage they were trying to produce, commenting on the action they were framing and using value words like "boring" and "interesting", suggests that the camera operators were experiencing the event in a relatively non-detached way while filming.

Design point: This recreational orientation that moves the focus of interest from the task (i.e. camera vision) to the recreational aspect of the activity (i.e. event vision) has wide implications for the user experience design of live media as it becomes more widely available. An 'amateur' production system should account for blending its users' event and camera vision. Thus the design of these productions should account not only for the users' visual focus of filming, but also for their actual engagement in the setting, whether it is an event as in this study, or other social situations or places that might be of interest to broadcast. This has been previously discussed with regard to other activities, such as tourism and hunting [3,13]. Given the pleasurable aspects of visiting places and events, the 'amateur' TV camera operator has to solve their task of filming and maintaining an awareness of the actions of other operators, in combination with event participation, and the devices' interface should reflect this. Designers therefore need to consider balancing the nature of the interaction away from requiring a high degree of concentrated attention to one in which they could segue into and out of collaborative filming. Such interfaces might support ad hoc planning and organisation, as this study points to a need for re-negotiating tasks and roles in 'amateur' settings. Thus, users filming an overview shot might request to replace close up camer operators when they were bored with their roles. What this requires is a communication system that is more a sophisticated than the vision mixer-oriented communication backchannel currently available on the IBS prototype. These extended functionalities could, for e.g. include unobtrusive signalling of an operator's level of engagement to all of the other team members, from full participation in the production to being on standby or even taking a break. Another way to support this would be to provide support for location or activity sharing, allowing users to check in to form ad hoc groups with friends or other people on location at an event, enabling low-effort social interaction and collaboration.

CONCLUSION

This paper aims to provide a new understanding of how live video can be produced as a group activity by non-professional users. Increasingly, advanced forms of internet-based user-driven production of video are emerging as part of

a larger trend that puts traditional notions of the 'amateur' and 'professional' into question. The IBS, as a live video production suite that leverages collaboration among empowered 'amateur' users to produce live video broadcasts, has allowed us to explore this novel design space. The combination of collaborative production and live transmission shows promise for new forms of participatory media, but also presents some serious challenges. This paper unpacks some of the challenges that this genre of production tools brings; some are related to skills, but lack of training alone does not appear to account for all of the issues that need to be taken into account when adapting such tools for a broader user group.

Our analysis shows that it is extremely hard for users to manage their 'loose' collaboration effectively to create topically relevant and meaningful footage for potential live broadcast that is visually complimentary to footage from the other camera operators–even when they are able to see the other camera operators and the final broadcast output. To a large extent, professional operators use their knowledge of the roles of the other people that they are working with, allowing them to anticipate their colleagues' actions and to align their own footage appropriately with them [21]. 'Amateur' operators cannot draw on the same knowledge and practices. Users must therefore rely on, and draw from their practical, everyday experiences as a resource for co-ordinating their own actions. We call such non-professional visual practices 'amateur vision' in homage to Chuck Goodwin, as a corollary to his term 'professional vision'. Although 'amateurs' cannot have full access to the rich set of practices and production skills of professional camera operators, as competent members of society, it is possible that they can replicate at least some of this from their everyday experiences with image media. In essence, they may be able to reverse engineer aspects of the production process. Yet, as we have shown, the inferences that they can draw from this are not necessarily useful. What also emerges from the analysis is that while taking on the task of acting as "proxy viewers" for others, users' camerawork practices are not detached from their roles as participants; they combine being a member of a camera team with experiencing the event individually as it happens. This is important in understanding the motivation behind this rather advanced form of collaboratively-generated content, as well as for the user experience design of production tools for non-professional live media. Looking towards user orientation in the recreational aspects of the activities that their work is embedded in is likely to be of increasing relevance for systems in which media content is produced by non-professionals, although this will be likely to differ according to the setting and activity.

REFERENCES

1. Bentley, F.R. & Groble, M. (2009) TuVista: Meeting the multimedia needs of mobile sports fans. Proc. ACM MM'09 (Beijing, China, October 19-24), 471-480.
2. Bergstrand, F. & Landgren, J. (2009) Using live video for information sharing in emergency response work. Int. Jnl of Emergency Management, 6(3/4), 295-301.
3. Brown, B. & Chalmers, M. (2005) Tourism and mobile technology. Proc. European Conference on Computer Supported Cooperative Work (ECSCW'03, 14-18 September, Helsinki, Finland). Springer-Verlag, 335-354.
4. Burgess, J. & Green J. (2010) The entrepeneurial Vlogger: participatory culture beyond the professional-amateur divide. In The Youtube Reader (Eds. Snickars & Vonderau). Library of Sweden, Stockholm.
5. Cisco Systems Inc. (2010) Visual Networking Index. http://www.cisco.com/en/US/netsol/ns827/networking_solutions_sub_solution.html#~forecast
6. Davis, M. (2003) Editing out video editing. IEEE Multimedia, 10(2), 54-64.
7. Engström, A., Juhlin, O., Perry, M. & Broth, M. (2010). Temporal hybridity: mixing live video footage with instant replay in real time. Proc ACM CHI, 1495-1504.
8. Figueiredo, M., Prado, P., Câmara, M., & Albuquerque, A. (2009) Empowering rural citizen journalism via web 2.0 technologies. Proc. ACM C&T, 77-84.
9. Garfinkel, H. (1964) Studies of the Routine Grounds of Everyday Activities. Social problems, 11(3), 225-250.
10. de la Flor, G. Luff, P., Jirotka, M. et al. 2010. The Case of the Disappearing Ox: Seeing Through Digital Images to Analysis of Ancient Texts. Proc. ACM CHI, 473-482.
11. Goodwin, C. (1994) Professional Vision. American Anthropologist, 96(3), 606-633.
12. Jayyusi, L. (1988) Towards a socio-logic of the film text, Semiotica, 68(3-4), 271-296.
13. Juhlin, O. & Weilenmann, A. (2008) Hunting for fun: solitude and attentiveness in collaboration. Proc. ACM CSCW, 57-66.
14. Juhlin, O. Engström, A. & Reponen, E. (2010) Mobile broadcasting-the whats and hows of live video as a social medium. Proc Mobile HCI'10 (Lisbon, Portugal), 35-44.
15. Keen, A. (2007) The cult of the amateur: how today's internet is killing our culture. Doubleday: New York.
16. Kirk, D., Sellen, A., Harper, R., and Wood, K. (2007) Understanding videowork. Proc. ACM CHI, 61-70.
17. Lange, P. (2008) Publicly private and privately public: Social networking on Youtube. Jnl. of CMC, 13, 361-380
18. Lehmuskallio, A. & Sarvas, R. (2008) Snapshot video: everyday photographers taking short video-clips. Proc. NordiCHI '08 (Lund, Sweden, Oct. 20-22), 257-265.
19. Macbeth, D. (1999) Glances, trances, and their relevance for a visual sociology. In Jalbert (Ed.) Media studies: ethnomethodological approaches. UPA:Lanham, 135-170
20. MacKay, W. E. (1999) Is paper safer? The role of paper flight strips in air traffic control. ACM ToCHI 6(4) 311-340.
21. Perry, M., Juhlin, O., Esbjörnsson, M. & Engström, A. (2009) Lean collaboration through video gestures: co-ordinating the production of live televised sport. Proc. ACM CHI, 2279-2288.
22. Shamma, D.A., Bastea-Forte, M., Joubert, N., & Liu, Y. (2008) Enhancing online personal connections through the synchronized sharing of online video. Proc. ACM CHI Extended abstracts, 2931-2936.
23. Shirky, C. (2008) Here comes everybody: The power of organizing without organizations. Penguin Press: NY.
24. Vihavainen, S., Mate, S., et al. (2011) We want more: Human-computer collaboration in mobile social video remixing of music concerts. Proc. ACM CHI, 287-294.