# Temporal Hybridity: Mixing Live Video Footage with Instant Replay in Real Time

**Arvid Engström[1], Oskar Juhlin[1], Mark Perry[2] and Mathias Broth[3]**
[1]Mobile Life@Interactive Institute, Sweden [2]Brunel University, UK [3]Linköping University, Sweden
{oskarj, arvid.engstrom}@tii.se mark.perry@brunel.ac.uk mathias.broth@liu.se

**ABSTRACT**
In this paper we explore the production of streaming media that involves live and recorded content. To examine this, we report on how the production practices and process are conducted through an empirical study of the production of live television, involving the use of live and non-live media under highly time critical conditions. In explaining how this process is managed both as an individual and collective activity, we develop the concept of temporal hybridity to explain the properties of these kinds of production system and show how temporally separated media are used, understood and coordinated. Our analysis is examined in the light of recent developments in computing technology and we present some design implications to support amateur video production.

**Author Keywords**
Video, editing, streaming, social interaction, television, media production, collaborative search, control room.

**ACM Classification Keywords**
H.5.3 Group and Organization Interfaces: H.5.3. CSCW; Synchronous Interaction.

**General Terms**
Human Factors

## INTRODUCTION

This paper examines a key feature in the production of live media: the co-ordination of events that are broadcast as they happen with recorded media of the event. This interleaving of current content with historical content is a common feature of live video and audio broadcasts, and may be used for a variety of reasons. Nevertheless, as a practical achievement conducted by multiple individuals involved in the media editing, content search and media generation, this is non-trivial: producing a meaningful account of the live event requires closely co-ordinated action by these actors. Examining the synchronization of live and non-live media offers implications for supporting the design of video editing technologies, a topic that has become of close concern to HCI and CSCW.

The relationship between video and interaction has been a longstanding concern within HCI. The discourse concerns problems in handling and gaining overview of volumes of digital video. It explores ways in which computing can be utilized to manage this complex data, e.g. through summarization [6,10,24] and browsing techniques [11,18]. There is also a parallel concern with live video and streaming video, in techniques for mediated talk [21,27], interaction with objects in the physical world through live video [25], remote video collaboration [16], and in work-place training and remote education [3].

On top of the research concerns, we see a growing interest in user generated video on the internet, which is now moving beyond sharing video files [5] to live streaming or live broadcasts from mobile devices (e.g. bambuser.com and qik.com). As bandwidth cost is decreasing faster than both processing power and storage cost [1], streaming, and managing streaming video is likely to become increasingly important; however, handling large video files and managing volumes of stored video data is likely to provide limitations on how this material can be practically made use of. This paper therefore extends HCI's concern with video to unpack the challenge of combining those two areas of research, i.e. handling of video files, and design for streaming video applications.

As a start, we turn to an ethnographic study and detailed interaction analysis [16] of professional TV production, to investigate *replay production*, a common feature of live broadcast video content. The purpose of this analysis is to unveil how time-critical video editing is at all possible, that is, how post editing (i.e. non-live) and live streaming is co-ordinated. We argue that unpacking how professional teams achieve this demanding practice can inspire the design process by making such narrative features available on emerging internet webcasts. Although replay editing is a common and cherished narrative feature in professional TV production, we do not yet see any use or availability of supporting tools on the internet. Following Perry *et al.* [22], we are not explicitly concerned with supporting the work of professional video producers, but the absence of non-professional users to study in this area and our interest in the *developed* coordination practices of skilled users has led us to examine professional production teams.

A distinctive feature of the practice of instant replay is that it is used to produce the salient extraction of activities as they emerge in the live action. In TV production, video replays are commonly used in broadcasts of live sports


*CHI 2010,* April 10–15, 2010, Atlanta, Georgia, USA.

events to enrich audio and visual images as they occur. Live television does more than simply showing events as and when they occur. Televisation allows the visual media to be edited and reviewed, and it does not simply provide a simulation of presence at the event for its remote audience. Different camera angles, close up images and a verbal commentary are commonly used to help viewers interpret and gain a sense of intimacy with the events unfolding and the actions of the players. In addition, viewers have come to expect that they will be able to review selected details of the often very fast moving action that contributes to important game play, such as goals, fouls, strategic mistakes, and so on, that are easy to miss on first viewing from the live broadcast footage. Replays are useful since they allow viewers to temporally return to a complex situation several times, from different viewpoints and at varying speeds.

The challenge in replay production emerges out of the necessity to juxtapose the edited historic video with live action in real time, and replay editing of streaming video is therefore severely *time-critical*. Replays are normally presented closely adjacent to an event in the live stream, and there is usually only a couple of seconds available for editing. Furthermore, the edited material has to be fitted into a broadcast where there is a continuing and somewhat unpredictable event going on. In performing their work, the replay editor has extensive live material available and can draw on *multiple video sources* that are logged continuously from the live event as it occurs.

We point to the way replay production draws upon the event itself (i.e. the sports game), with its inherent game time structures, with intermittent action and breaks. Here, the TV production's dependency on game time (which it cannot control), is used as a contextual resource in the replay production To illustrate this, we identify several resources and practices through which the whole TV production team collaborates to make possible time-critical search and editing of logged video data. We show how video material with different temporal trajectories is split up on different media (audio and video) or on different screens, and how this is useful in editing and search. We argue that the use of replay features in user-generated live video might fit with similar events that are organized and narrated, such as games or performances, that may occur in the physical world or online [7].

## RELATED WORK

Research on the relationship between video and interaction has been ongoing for the past 20 years. A main approach has been to aid users, both professional and amateur, in reviewing and accessing volumes of video content, through various summarization techniques [e.g. 6,10,24]. More hands-on approaches assist users in the editing process through novel ways of browsing and visualising video [9,11,18,26], while Kirk *et al* [17] take a more user-centric view of video to inform the design of editing tools.

Sports video has itself been widely studied and has been the focus of technology research developments and commercial innovation in recent years. Interest in the topic has been motivated by both its large, and commercially significant, audience numbers, and technically, through widespread dissemination of enabling technologies such as fast and large capacity hard drives, high-speed digital cameras and real-time computational image processing. Replay itself has a relatively short history in live television, with a beginning in sport productions from the 1970s, where replay operators worked with analogue tape machines [28]. Although digital replay technology is now used, it is still relatively expensive and technically challenging, but it is increasingly being used in televised sports and other live genres.

As in the more general HCI literature, summarization of recorded media forms the dominant topic of technical interest. Major application areas of computer-based research into instant replay include tactics analysis, tracking of objects and landmarks, automatic highlight extraction and various forms of event and landmark detection using image analysis [29]. A number of frameworks for event analysis and highlight detection from video have been proposed, based on audio [23], camera work [31], heuristic analysis [12] and combinations of visual cues, motion and audio [23,29] to support image detection and selection. Some of these systems are designed to specifically recognize replay sequences in the finished broadcast, using them as indicators of importance in highlight extraction [2,31]. Other recent efforts in image recognition [30] provide more nuanced accounts of replay insertion into live broadcasts, touching on the role of skilled subjective decision-making and narrative concerns in video production, and attempting to put these issues of automation in image selection into more realistic contexts of use. However, as these studies have only examined the finished product, i.e. the broadcast program of live sporting events, as their raw data for the analysis of the production process, they are unable to access the ongoing process of narrative production and the skilled practices of sense-making that the production teams undertake when working with the video material.

Multicamera work and vision mixing, as it is done in studio productions and live sports television, has also been examined. Engstrom et al [13] present a mobile broadcast system for amateurs, and Mondada [20] presents a detailed analysis of camera selections and 'split screen' use in the editing process of broadcast content by production directors to dramatize moments of conflict. In terms of collaborative production practices, camerawork and camera selection has been analysed in detail and described in terms of proposal-acceptance [8] in the studio and the use of indexical gestures [22] in live sport by camera operators to 'point' to action and to demonstrate their availability for selection. In the live sports setting, this is further complicated by rapid movement and more unpredictable live action, whilst the developing action needs to be covered from multiple angles so as to produce a coherent experience for the viewer.

## PRODUCING LIVE TV SPORT AND INSTANT REPLAY

The main direction and visual production of a TV show is conducted in the production control room. The setup of the

room contains a 'gallery' of video monitors displaying all camera sources centered around a main broadcast monitor and a preview monitor. An intercom system enables communication between this room, the camera operators and adjacent production units.

A fundamental difference between the production of pre-recorded and live television is that the former splits the production process into two phases (filming on location and editing afterwards), while live TV is simultaneously recorded and transmitted. This is achieved by operating multiple cameras and mixing their live content together with recorded video sequences ('instant replay'), audio and supporting graphics. The ability to create instant replay material in the production of contemporary television relies on the use of non-linear (or tapeless) media, which allows 'random access' to stored video footage. Video and audio materials are captured to a storage device, which allows recorded footage to be searched, segmented, resequenced and played back. In live sport that involves the use of multi-camera recordings, these systems allow program editors to cut into the live broadcast to show recorded footage from cameras that were not initially selected for broadcast, allowing the use of multiple angles on actions taking place during the game and at different playback speeds. The role of the instant replay operator is to act as an editor, assessing and selecting sequences very rapidly as soon as they occur to create material that can be cut into the live footage when possible or appropriate. These operators are not just technical operators, skilled at working with the video to produce content when requested – they need to be highly attentive to the developing game in producing relevant and timely footage: quoting a live sports producer, 'You've got to give the tape operators a lot of leeway. And they've got to know the sport. They really have to know it.' [28, p. 65].

**Figure 1. Video gallery with replay screens in front**

## METHOD AND SETTING

The data collection took place in the outside broadcast (OB) studio located in a custom-fitted bus outside an ice hockey arena. The major part of this studio is taken up by workstations for the vision mixer (VM), the producer, the script and the graphics operator, all facing a video gallery (Fig. 1). This video gallery displays all the visual resources the VM has at hand; manned and unmanned cameras placed around the arena, two monitors showing the replay operator's work and one display for graphics overlays. Behind the main video gallery, but in the same space, is the replay operator's workstation, described in detail below. The VM and the replay operator (RO) can communicate verbally and hear the commentators on loudspeakers inside the studio. The VM is directly audible to the commentators via an intercom headset, while the RO can speak back to the commentators by pressing a button to activate the intercom. In all, a team of about 16 people, including camera operators, sound and image engineers collaborate to produce the live broadcast.

Data collection on the live TV production process involved a number of sources and participants, and took place during two ice hockey matches in February 2009 in Sweden. The majority of the empirical data collected and presented in this paper has involved ethnographic observations and video recording within the OB studio, and the analysis presented below relies on this whole empirical corpus. In addition to the observational data, we conducted interviews with two replay operators, the first of whom gave us a course in its operation. This course was video recorded and reviewed, allowing us to become acquainted with using the system. In the interview with the second replay operator, we discussed our understanding of the system and its use practices to check on our understanding, and to examine different perspectives on their use. Whilst we recognize that interviewing two replay operators does not provide a substantial corpus of data, this is a highly specialized job, with very few professional operators.

In addition to formal interviews, we engaged in numerous ethnographic interviews with various production teams on the use and practices involved in instant replay and its incorporation in live broadcasts. This interview data itself is not reported in this paper, and, in line with our use of interaction analysis [16], the analysis reported here relies on the observed actions alone. Our interest in the selection of material for analysis primarily focused on key events prior to, during and directly following the selection of replays for broadcast. 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 three cameras, this resulted in over 15 hours of tape recordings (including pre-and post match events). One camera was directed at the monitors in the control room, whilst the second framed the replay operator's hands from the side, and the third overlooked the operator's face and the screens in front of him. All participants agreed to participate in the data collection and we have accorded them anonymity. The recordings have been repeatedly viewed in team analysis sessions, and core events transcribed and categorized. The material presented has been translated from the original Swedish into English.

### Technologies in play

The technology used by the replay operator in the data examined in this paper was the EVS Multicam LSM (Live Slow Motion) hardware, which allows recording and manipulation of non-linear media recordings. This EVS unit is coupled with an XT[2] production server; visually, this drives a monitor featuring a split screen display of four live camera feeds and two additional monitors for viewing the

replay operation (Fig. 2). This setup records multiple live camera feeds to the server continuously throughout the game, and enables the operator to go back in time to any of the camera feeds, search within the video and edit short sequences to be replayed. Events in the material can be accessed instantly as they occur, and individual sequences can be edited into playlists to provide multiple camera angles on a situation [32]. At this point, and in the same way as with the live cameras, these replay image sequences can be selected and cut into the broadcast feed by the vision mixer. Although the VM controls the feed selected for broadcast, within the live replay, the EVS operator maintains control of playback, such as where the replay begins from, its playback speed, and any cuts between cameras within the recorded material.

**Figure 2. Replay operator's monitors: 4-window split screen on top and two working screens below**

The XT[2] control unit can be seen in figure 3. The labelled elements correspond to the following key functionalities and their corresponding interface controls: 1) camera selection interface allowing the operator to select from up to six synchronized live camera feeds; 2) video jog wheel which is used for searching within the video stored on the server; 3) playback control lever (from -400 to +400% of normal playback rate); 4) a video bank for storing clips for later access, individually or as playlists.


**Figure 3. EVS Multicam Live Slow Motion Harware XT (2)**

The replay operator's work involved the continuous identification of potentially interesting situations in the game. When such a situation took place, he typically went through the cameras to examine which held a suitable framing of the situation by rewinding the video that had just been stored on the server. He would then select one (or more) video streams that showed this situation. On locating this he would set an 'in-point' to the selected feed and then typically waited for directions from the vision mixer. If the vision mixer, who relies on the replay operator to have done just this, calls "EVS…", the operator prepares to roll the sequence upon the command "…now". If no such call is made, the sequence is stored in a video bank.

As the game unfolds, the operator keeps track of multiple situations at the same time, editing playlists of sequences and searching for material further back in the video feeds. For these purposes, he has two working channels displayed on a "program" and a "preview" monitor, as well as controls for customization and playlist management.

## ANALYSIS

In the following, we analyze the interplay between video editing and live streaming i.e. replay production, with a specific concern on how editing tasks, such as searching and archiving video snippets, are managed alongside the on-going live broadcast.

Replays hold an important role in the narrative structure of a TV production, although they only form a part of the production and have to be coordinated with the rest of the material that compose the mediated event as it is broadcast. We will bring forth a number of activities that both frame this work, and at the same time are used as resources in replay production. These activities include the constantly available sound and visible actions of the players and crowd at the event, the sound of the commentators talking about on-going events, their skilled readings of previous and possible future game play, the structure of game play with its temporal pattern of intermittent play action and pauses, the use of narrative structure in camera selection, the mix of live footage and replay, and the practices and demands of the production process. The excerpt below shows how the replay operator (RO) draws on the available imagery on his video screens, but also on many other resources in the environment. To increase the readability of this complex activity, we provide an analysis with reference to a single replay situation. To illustrate this, we provide a schematic description of the situation (fig. 4) with reference to clock time. In the diagram, we separate timelines that correspond to the game events unfolding on the rink, the broadcast audio from the commentators, the actions of the replay operator (what he says, his visual focus as determined by his orientation to resources, and the tasks that he is currently working on using the EVS equipment). At the bottom, the image selected for broadcast by the RO is shown. The arrowed lines denote key references between different parts of the timeline. For the analysis, the grey area forms perhaps the most important feature of the replay production, and covers the hybrid temporal space that the operator is working within and cross-referencing, i.e. current and historical events.

### Looking back and looking around

There is an expectation that pre-recorded material can be brought into play to either help explain the current situation, or to fill in a relatively uneventful break in the action when

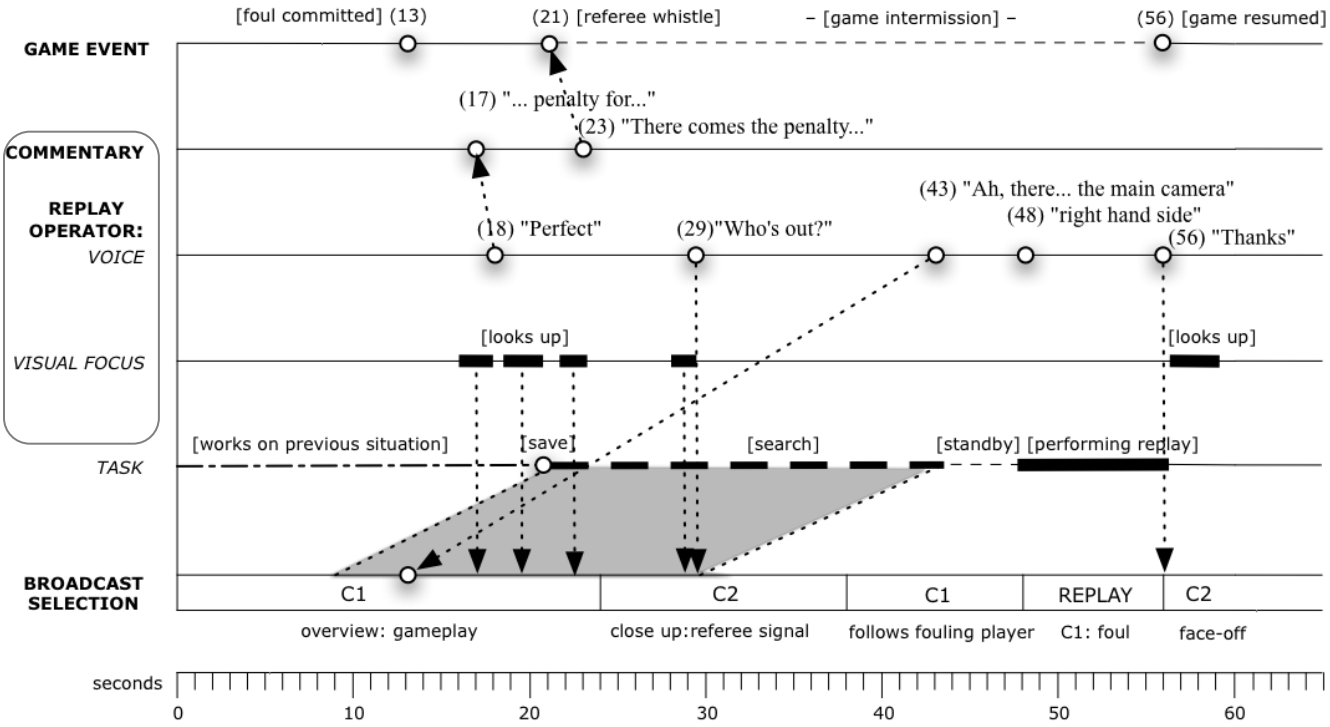


**Figure 4. Schematic outlay of time line for activities**

there is a pause in the game, that is when it is 'out of play', for whatever reason. As soon as the replay operator decides to engage with a particular replay sequence, there is a danger that new material might emerge which also demands his attention. He will therefore find himself in a situation where he still needs to attend to both the unfolding live action and the constant flow of pre-recorded video material for editing. There is no way in which he can pull away from one to do the other, as they are both necessary aspects of his work that require his attention.

We discuss how this is achieved by analyzing a section where the replay operator is editing a recording when a new situation emerges in the live game. This new situation requires him to engage in searching through his video bank for relevant recorded footage. Such searching for a particular situation from within the video bank is a common and critical activity in the production of replay footage. In the following, we will unpack how this is conducted as an ongoing practical achievement, using data from an event that relates to a penalty call that a referee has made; penalties are a type of event where the operator is normally expected to provide a replay of the situation. An audio transcription of the event is listed below (timing information is shown in brackets, allowing cross-reference to the activity timeline in fig 4). Just prior to this sequence, the RO has been editing recorded footage from a previous event, but 13 seconds into the sequence, a player commits a foul (figure 4):

1. Producer: damn, that's a short period.
2. Commentator: excellent of Huddinge [pause] (17) penalty for
3. RO: (18) [looks up] perfect.
4. VM: period over. Yeah didn't even get out the live odds
5. Commentator: Huddinge is pressing all right here, Karlberg away with his stick [whistle blows]. (23) There comes the penalty with…

This foul is visible on the screens that present the live action, but the replay operator looks at the logged video data. At this point, his attention is on the replay screens and control access interface where his clips are saved for later broadcast. When the commentator says there is a penalty (line 2) he briefly looks up at the live screens. In ice hockey, when the referee observes a foul, offenders are normally sent off from the rink to a booth for a two minute enforced break from the game. When the referee blows the whistle and the commentator again says "penalty" (line 5), he looks up again before returning to editing the previous section. So, he looks up at the screens showing the broadcasts at both situations at the point when the commentator announces the penalty (lines 2 and 5), but manages to attend to both his video bank and this live action. It would thus seem that he uses the commentators' description of the game in the voiceover, which can be heard over the loudspeakers as a behavioral trigger to get awareness of the visual content of the live video feeds.

It is the availability of several media, visual and auditory, that makes this split of attention possible. A pair of loudspeakers in this small room equipped with dozens of screens of various sizes plays a central role in getting the job done. The visual screens present various temporal trajectories [cf. 4]: they display action in real time and in past time, but these are complemented by the audio, which represents only what is going on in real time.

6. VM: two, there
7. VM: number 1, number 1, penalty

8. Expert commentator: I think Huddinge is working on rather well
9. Producer: now there's a game.
10. Expert commentator: I think that Huddinge has very good
11. RO: (29) Who is out?

The replay operator starts to work on the penalty directly after finishing previous tasks. He pushes some of the buttons on the top row, which indicates that he is storing previous work. He then starts to work with the new penalty. He rewinds the footage from camera 2 and reviews the event that took place in front of the goalmouth. He starts from the referee's call and works backward in the video from this. He appears not to find anything of relevance and selects camera 1. He continues to rewind further back, and finds a situation where two players are very close to each other. He then runs that situation back and forth once on his screen. He then rewinds further backwards until he finds another situation in which two players from the opposing teams are physically close to each other at the rink side. Then he starts moving the footage forward again and says loudly "who is out?" (line 11). So, he not only switches in between different cameras, to get new angles, but also varies the playback rate of the replay for detailed investigation into the video feeds. Furthermore, he also uses the visible proximity of players from camera 1 as a resource for his search, which is an obvious requirement for this offense being committed.

The RO's problem in searching for the offense (occurring at 13s) is that the commentators' reference to it (occurring at 17s, figure 4) is not temporally synchronized with the penalty event on camera. The commentators' reference is to the referee's decision rather than to the offense per se. Thus, the cameras do not display the foul when the commentators announce it, and when he looked up (figure 4), and he does not really know what he is looking for when he starts searching, hence posing the question "who is out?" (line 11). This request for information makes his task for finding the offender publicly available to the other people who can hear this. When he cannot find the situation by a quick glance through the video banks he can ask for more information on what he should be looking for.

12. VM: (30) who's out? Check that
13. Expert commentator: Västerås has 2 0 but Huddinge has had the latter part as I think
14. VM: (32) hooking

The VM quickly passes on the question for others to look (line 12) and then adds that this involved "hooking" (line 14), i.e. the referee's categorization of the situation. The referee's interpretation and his hand signal was framed by camera 2 and selected by the vision mixer for broadcast. This indexical reply by camera 2 to the VM (i.e. the camera pointing towards the offender) also follows a standard procedure that occurs adjacent to penalties, where one camera operator is under instruction to follow the referee, while another camera operator selects the offender. The access to different visual streams also gives the replay operator additional resources to search for the offense:

15. Expert commentator: the puck slides much more on the right side of the rink
16. VM: on someone from Västerås
17. VM: One, there
18. Expert commentator: skates in a good way
19. VM: name now
20. Expert commentator: has to adjust the aim
21. VM: Go!
22. Expert commentator: we saw it earlier
23. VM: eighty [as the player number becomes visible in camera 1]

This access to different visual streams makes it possible for the VM to see the number of the player ("80") committing the offense and he can answer the question put up by the replay operator (from line 11).

Turning to the ongoing broadcast (bottom line in figure 4) and looking through the whole sequence again, we can see that the vision mixer selects camera two (line 6), when the referee blows the whistle. When he says "number 1, number 1, penalty" (line 7) this is a renegotiation of a commonly used format [22]: normally, camera 3 would have provided such close up shots. But this camera operator had recently left his position for another task and was currently unavailable. Camera 1, which normally provides overview shots, was ordered by the vision mixer to take on camera 3's task and frame the offender or the "penalty" (line 7). Camera 1 is then selected for broadcast (line 17). In this case, collaboration by the production team facilitated the replay operator by giving him time to search for the offense. Working together, they produce a narrative of the game, including specific topics and framing by the cameramen to convey the ongoing action in the game to the audience. However, in doing it in this particular way, they also buy the replay operator time to do the search, as well as to narratively link the potentially upcoming replay with the live event. For example, after a penalty, the camera operators who are assigned to provide detail shots (decided by prior agreement) are told to alternate between close up shots of the referee, the offender and the penalty booth, in that order. Thus, the camera operators do several jobs. They provide broadcast material to the vision mixer, but also make time for the replay operator to perform his search.

24. VM: Go!
25. Expert commentator: Ahlström shot above earlier
26. RO: (43) yeah there [quietly]
27. VM: you got it

Finally, the operator locates the offense (line 26). His utterance "yeah there" (line 26) could be interpreted a comment to himself, but it is also made it aloud to the people around him, showing that he has found what he was looking for. This exclamation of success is socially relevant, because his search has been made public to the rest of the production team through the VM's call for help (line 12), and his finding the solution closes this task. This identification by the operator is also potentially important for the VM, who now has an opportunity to replay the offense. The VM then asks for additional clarification from the RO "you got it" (line 27) before selecting the replay for broadcast.

28. Expert commentator: so it is about
29. RO: (45) the main camera.
30. VM: yes, EVS now
31. Expert commentator: getting the puck towards the cage and create…
32. Commentator: they are creating some…
33. VM: now!
34. RO: (48) right hand corner [decelerates playback speed]
35. RO: (59) right hand corner, lads [returns playback speed to normal]
36. Expert commentator: rebounds
37. Expert commentator: yeaah, I think so
38. VM: yes, there
39. VM: what's he doing then?
40. Commentator: there we have the penalty on Holmgren
41. Commentator: not much to say about that

At this point, the RO now instructs the commentators as to where they should look on the screen when the replay goes live on air. The commentators do not have the opportunity to go back and check through the video bank. They get only one opportunity to comment on the replay broadcast. The replay operator utilizes the opportunity to talk directly to them, through pressing a button opening an audio link to their headsets when he says "right hand corner, lads" (line 35). This provides the commentators with information about where on the screen the event will occur. It also shows that this is unusual footage in the sense that the shot itself does not make the action of interest particularly salient. So, the commentators could need a hand to see this event. The VM overhears the conversation and acknowledges that he also saw the offense (line 38). The RO also decelerates the playback speed just as the hooking occurs, which makes it easier to locate the foul. The commentators are now able to claim that the referee made a correct call (lines 40) and that there is "not much to say about that" (line 41). Thus, the replay operator can randomly access any point in game time, review it and reproduce it to the commentators, coordinating his ability to work across time with the others' ability to access only real time content.

As we have shown, searching through the video bank can be a highly collaborative task in which the RO, VM, commentators and camera operators work together to solve the task of creating broadcastable live *and* non-live footage. In this case, the commentators' reference to a previous action that was deemed relevant for search and broadcast did not occur temporally adjacent to that action (i.e. they did not talk about the offense when it was live), but rather they talked about it only after they had seen the referee's arm signal. Thus, even though the replay operator had access to the live feeds, he appears not to have seen or recalled the sequential unfolding of this offense and precisely when it had occurred. By mobilizing his technical and social resources in the production team, the RO managed to discover the reason, or at least the sequence of actions, that led to the referee making this penalty call and find an appropriate edit point for the replay.

### Summary of resources

The detailed analysis above reveals some of the ways that replay operators act on logged material and at the same time engage in the live broadcast, as well as how they utilize socio-technical resources other than their video screens during severe time constraints. Below, we pull these out to highlight the various mechanisms of how this is achieved.

**Temporal coordination through media threading:** at the same time as searching through logged data, the replay operator listens to the on-going audio commentary, using this as a resource to check the live video feeds on occasions where they talk about possible replayable topics. These occasions function as situated audio "tags" that allow work on the live video streams to be synchronized with pre-recorded media.

**Tracing historical references backwards in time:** the RO used the live camera's image of the referee's arm signals as indicators of the actions that had occurred previously in the encounter between the players, i.e. one player illegally hooking the other player with his stick. He then used the other camera's view on a player skating towards the penalty box to draw the conclusion that this must be the offender. Thus, as he was searching through the video materials, he continuously obtained more visual information from the live feeds to help him make sense of the logged media. Here, the referee's signal, the commentators' remarks and the camera operator's selection of a player skating towards the penalty box were references that helped make sense of previous actions. It follows that the operator not just acts on what he sees, but reads this as references of what happened previously. In this case, the replay operator 'back projected', or 'retrospectively indexed' (cf. Goodwin's [15: 384] 'prospective indexicals'), what he saw live onto what had gone before to identify the offender. The commentators' discussion on what had happened previously is also used a resource in this respect. Importantly, both commentators and the VM switch focus between referring to what is currently happening, and to the past broadcast.

**Distributed and parallel search:** Verbalizing the ROs search out loud allowed the production team to search for an important event for replay simultaneously, and thus cover more material in the brief time available. Here, the VM identified the player skating towards the penalty box while the RO was scanning back and forth in his video bank. We see here a form of "functional separation" of the search between the VM and the RO, bringing several more screens into use for replay production than could be used by the RO alone.

**Synchronising production with game time**: Replay production is oriented towards game time in that it allows the production team to fill gaps in game play. The intermittent structure of game time, and especially the pauses in play, provides opportunities to focus more on editing and less on the live action. This is because it is unlikely that any new game action will emerge that is appropriate to use for replays during this time. In this case, the VM selects cameras 1 and 2, framing the referee and the offending player skating towards the penalty box.

**Narrative formats supporting replay production:** The live feed of video provided by the camera operators during game intermissions is helpful for the RO, even though he is not using this material in his edited version. It is useful because the narrative format changes when the referee blows his whistle. At this point, the camera operators switch from following the game action to showing what had happened. In this case they switched from *following the puck* to trying to *providing an account,* by exposing who did the offense earlier on. This switch in narrative formats has two consequences for the replay producer. First, it provides the RO with time to search and edit his material. Second, it provides him with a bridge in the narration of the game in between the actual situation and the replay of it.

The live production and the replay production are intricately meshed at various levels. Replay production is possible not just because the operator can rewind his tapes, but because the whole production apparatus can be organized to either work in synchronization with what is going on, or to display, refer or comment on what had happened previously. The availability of game structures and narrative structures are important in mediating the combination of those two forms of temporal content.

## DISCUSSION AND IMPLICATIONS FOR DESIGN

This study explored editing techniques and live broadcast production from the perspective of research in media and interaction design, and have unpacked some of the interactional mechanisms through which replay features was made possible. In the following, we expand on these findings picking out features of relevance for the design of computer-based systems.

**Temporal hybridity:** Replay production is very much about working with and on time-lines. We suggest that it displays a sort of temporal trajectory [cf 4], which has not yet been accounted for in the discussion of temporality and narrativity. In mainstream TV, broadcasts are produced to provide a shared experience for the audience. From the viewer's perspective, a televised hockey game consists of shots of live action followed by shorts sections of historic video data showing recent actions, before cutting back to the live action. In this sense, the set up differs from the temporal trajectories discussed by Benford *et al.* [4] that depend on allowing both narrative time and consumption time to vary. However, in television, the final outcome should be designed to allow users to consume it together and at the same time, which is the quintessence of live TV.

In the analysis, we show the temporal arrangements in the production activities, and how temporally disjointed narrative material is *presented* to the TV audience. Here, historic, or logged, video data is co-aligned with on-going action. We characterize this work as involving *temporal hybridity*, adapting use of the media term "hybridity" from Manovich [20]. In our case, temporal hybridity denotes how the real time co-alignment of media types *with various temporality is co-aligned, at the same time within the production environment*. This occurs in two ways: first, in the hybrid combination of live audio with historic video data, and second, in the video material displaying both live action and historic action. The concept of temporal hybridity makes visible a somewhat odd place in media production. A concept to denote temporality management is a resource when further considering temporal aspects of media production and consumption. Working to create alignment within this set of hybrid forms is essential for succeeding in achieving broadcasts where mediated events take place that need to be accounted for in real time.

**Extraction and summarization of video content:** We now turn to consider managing large volumes of video data, and in particular efforts to automate the extraction and summarization of video content. We have argued earlier that there has been a shift toward live video in both research and video usage, enabled by emerging technologies, and that the intersection of these two areas requires new forms of technology solutions to deal with this. Nearly all of the summarization solutions presented to date apply computing to the finished broadcast product post-event, to provide more sophisticated control of, or extract highlights from *non-live* material. As the majority of this research is not scoped towards real-time production, we can only discuss the challenges and possible benefits of a shift towards live video in this field in relation to some early attempts at real-time automation and the ideas they represent.

Although increasingly powerful image recognition certainly holds promise for making media production more effortless [28], examining the specifics of manual practices such as camera selection [22] and replay production problematizes these claims. Here, we have shown how replay production draws on a range of socially mediated resources and multiple temporalities in order to form a solid comprehension of events that is necessary for the meaningful production of live broadcasts. These resources and domain specific knowledge are used to make informed decisions on what actions to replay, when to cut, and how best to render action through a selection of camera angles and control of playback speed. This study therefore presents three interesting challenges not accounted for in entirety in any automated systems to date:

1. Temporal hybridity in search. The operator combines domain knowledge with search in multiple media and access to layers of historic time to find replay clips.
2. Uncertain and limited time constraints. The operator has to work within the available, although unpredictable, window of broadcast time to search for, produce and "perform" replay sequences.
3. Aesthetics and accountability. The operator has to work within the constraints and values of the TV production process, using multiple camera angles and control of playback speed to render a narratively meaningful and error free replay for the audience.

The promise computing holds in streamlining and lowering the barriers of producing live television in general is still valid. But given these challenges, and the fact that replay

operation commonly is managed by a single operator, the notion that automating this particular function would "substantially cut down the crew size" of television teams [30] seems exaggerated. Furthermore, we have shown how the replay operator's search work serves an important secondary role supporting all key production roles, from vision mixer to commentators, in informing their narration of the live event. Automation may only provide a partial solution to this problem of search and image insertion into the live broadcast. What may be more useful here is to place automated 'bookmarks' derived from image analysis into the live footage to assist search and retrieval.

**Distributing replay production:** The step from professional instant replay production to its amateur production is likely to be greater than for vision mixing [22], since it depends to a much larger extent on technical dexterity and specialized skills. It is unlikely that straightforward inclusion of replay features in broadcast systems will result in amateur productions of a quality that matches anything like professional media. However, there might be a means of transforming replay production activities to make them less cumbersome. We have already commented on the unsuitability of automating replay, but it might be possible to collaboratively distribute the replay production process. We envision a set of tools that further distribute instant replay. As we have shown, the RO currently attends to voice commentaries in real-time editing. Yet these links between the verbal remarks and the visual content are only available momentarily: replay might benefit from formal annotations, making such tags permanently available. It might be possible to involve camera operators, commentators and the VM in tagging live video. As they pursue a task at hand, they could suggest or select possible edit points and content definitions to perform live tagging of video streams. The replay operator would then receive pre-structured video data. Pre-tagged live video would buy the RO time to focus on other time-critical activities than searching for content, and could potentially support useful input for simple automatic live-to-recorded editing transitions in the future. This could be useful both in professional production in situations where the video data is overwhelming, as well as in amateur production.

**Talk in temporal synchronisation:** The study reveals the importance of talk as a means of coordinating replay production, and specifically, its role in supporting the alignment of temporally separated media. Although the replay operator was surrounded by a large number of screens, that data shows how he often acted on the sound of the commentators' conversation. Furthermore, there was an on-going conversation between him and the VM. Interestingly, this differs from the communication between the VM and camera operators. As shown in other studies [8,13,22], the VM talks to the camera operators ('camera 2, now!'), whereas they can only respond by framing interesting topics with their cameras. Replay production is different in that it depends on bi-directional communication, and we suggest that future replay production systems should explicitly include resources for coordination work through talk.

**Mapping replay production to game time:** We have shown how replay production is both constrained by the game structure, but also uses it as a resource. It is constrained by game time because there is a requirement to deliver a replay close in time to a live event, yet game play cannot be controlled or manipulated by the TV team. However, the game structure also helps them in providing the replay: when the referee blows the whistle and the game clock is halted, it is very unlikely that new situations will occur that will be available for future replays. This has two consequences. First, the RO can turn all his attention to editing. Second, the rest of the team (VM and camera operators) can change their focus from following the action to accounting for what just happened. In this sense, their action can become a contextual resource for the RO, for example, in searching through the logged data. In future distributed production systems where the RO is not proximate to the other members of the production crew (e.g. where video production is carried out on mobile devices), it may therefore be useful to provide some form of indication to the RO that these other members of the production team are not currently highly active and can be co-opted into activities to support them, such as we have seen in search.

**Expanding replay to other high intensity activities**: We argue that our findings are important for considering what types of other broadcast situations could benefit from video replay. We suggest that such activities will support replay production if they have some sort of intermittent temporal structures such as action and pauses. Narrated events of various kinds usually have similar features, i.e. a rhythm between tension and relief, such as events or performances of various kinds or game play. In general, these types of events also share a basic feature with ice hockey in that they are in some sense designed for an audience. They provide a somewhat saturated game experience during a limited period of time, and the use of replay is also of relevance due to this saturated experience. Ice hockey provides fast action, and it is visually demanding to understand what is going on. Replay production can therefore be seen as a resource to unpack an activity designed for high intensity and complexity. Considering extending the use of replay might therefore be most rewarding for video recording complex and intense activities, and less useful for areas such as decision-making, learning and communication. The latter activities are often oriented towards the long term and are less complex than, for e.g., a musical performance or pervasive gaming.

## CONCLUSION

The longstanding interest from the CSCW community in video combined with recent HCI interest in user-generated content, video posting sites, and mobile reporting has led to an increase in HCI research on video. Much of this research seems to assume that we can produce simple tools that ape the professional ones. However, in reality, professional live broadcast is highly skilled and intense, and this analysis should encourage a more realistic appreciation of the task for amateur production.

Instant replays are a key media feature in rich live video content. Today, the production of such media is in the hands of skilled professional TV-teams. In this paper we detailed empirical findings to unpack how they manage to organize themselves to search, edit and broadcast previously recorded video closely adjacent to live broadcasts. Our intention has been to inspire the design of similar tools to support amateur and less professional media producers, in order to support more advanced user-generated live broadcasts that can be viewed when mobile or on the web. Furthermore, we suggest that our focus on the juxtaposition of media with different temporal trajectories, or what we have termed as temporally hybrid media, might inspire new kind of narratives both in video productions and games development.

## ACKNOWLEDGMENT

The research was partly made possible by a grant from the Swedish Governmental Agency for Innovation Systems to the Mobile Life VinnExcellence Center, in partnership with SonyEricsson, Ericsson, Microsoft Research, Nokia Research, TeliaSonera and the City of Stockholm.

## REFERENCES

1. Anderson, F., Free: the future of a radical price. Hyperion Books, 2009.
2. Babaguchi, N., Kawai, Y., Yukinobu, Y., Tadahiro, K., Linking live and replay scenes in broadcasted sports video. In Proc. Multimedia'00, 205-208.
3. Bargeron D., Gupta A., Grudin J., et al., Annotations for streaming video on the web. ACM CHI'99 extended abstracts, pp. 278-279.
4. Benford, S. and Giannachi, G., Temporal trajectories in shared interactive narratives, In Proc ACM CHI'08, pp 73-82.
5. Blythe M., Cairns P., Critical methods and user generated content: the iPhone on YouTube. In Proc ACM CHI'09, pp 1467-1476.
6. Boreczky J., Girgensohn A., Golovchinsky G., *et al.*. An interactive comic book presentation for exploring video. In Proc. ACM CHI 2000, pp 185-192.
7. Bowers, J. (2001) Crossing the line: a field study of inhabited television. BIT, 20(2), 127-140.
8. Broth, M. The Production of a live TV-interview through mediated interaction. In Proc. Int. Conference on Logic and Methodology'04. SISWO, Amsterdam.
9. Casares J., Long A.C., Myers B., et al. Simplifying video editing with SILVER. In ACM CHI'02 extended abstracts, pp 672-673.
10. Christel, M., Smith, M. A., Taylor, C., Evolving video skims into useful multimedia abstractions. In Proc. ACM CHI'98, pp 171-178.
11. Dragicevic, P., Ramos G., Bibliowitcz, J., Video browsing by direct manipulation, In Proc ACM CHI 2008, pp 237-246.
12. Ekin, A., Tekalp, A., Automatic soccer video analysis and summarization. IEEE Trans. Image Processing, 12, 7 (2003), 796-807.
13. Engstrom, A., Esbjornsson, M., Juhlin, O., Mobile collaborative live video mixing. In Proc. ACM MobileHCI'08, 157-166.
14. Fussell S. R., Setlock L. D., Kraut R. E., Effects of head-mounted and scene-oriented video systems on remote collaboration on physical tasks. In Proc ACM CHI'03, pp 513-520.
15. Goodwin, C., Transparent vision. In: Ochs, E., Schegloff, E.A., Thompson, S.A. (Eds.), Interaction and Grammar. CUP, Cambridge, pp. 370-404, 1996.
16. Jordan, B., Henderson, A. Interaction Analysis: Foundations and Practice. Journal of the Learning Sciences, 4, 1, (1995), 39-103.
17. Kirk, D., Sellen, A., Harper, R., Wood, K., Understanding videowork. In Proc. ACM CHI'07, 61-70.
18. Li F., Gupta A., Sanocki E., Browsing digital video. In Proc. ACM CHI'00, pp 169-176
19. Manovich L., Understanding hybrid media, in Hertz (Ed.), Animated Paintings. San Diego: San Diego Museum of Art, 2007.
20. Mondada, L. Video recording practices and the reflexive constitution of the interactional order: some systematic uses of the split-screen technique. Human Studies, 32, 1 (2009), 67-99.
21. O'Hara K., Black A., Lipson M., Everyday practices with mobile video telephony, In Proc ACM CHI'06, pp 871-880.
22. Perry, M., Juhlin, O., Esbjörnsson, M., Engström, A. Lean collaboration through video gestures: co-ordinating the production of live televised sport. In Proc. ACM CHI'09, 2279-2288.
23. Rui, Y., Gupta, A., Acero, A. Automatically extracting highlights for TV baseball programs. In Proc. ACM Int. Conference on Multimedia (2000), 105-115.
24. Song, Y., Marchionini G., Effects of audio and visual surrogates for making sense of digital video. In Proc. ACM CHI'07, pp 867-876
25. Tani M., Yamaashi K., Tanikoshi K., *et al.*, Object-oriented video: interaction with real-world objects through live video. In Proc ACM CHI'92, pp 593-598.
26. Ueda H., Miyatake T., Sumino S., *et al.*, Automatic structure visualization for video editing. In Proc INTERCHI'93, pp 137-141.
27. Veinott E. S., Olson J., Olson G.M. *et al.*, Video helps remote work: speakers who need to negotiate common ground benefit from seeing each other. In Proc ACM CHI'99, pp 302-309.
28. Verna, T., Live TV: an inside look at directing and producing. Focal Press: London, 1987.
29. Wang, J.R, Parameswaran, N., Survey of sports video analysis: research issues and applications. In Proc. VIP'03, Austalian Computer Society, Inc (2004), 87-90.
30. Wang, J., Changsheng, X., Engsiong, C., Kongwah, W., Qi, T., Automatic replay generation for soccer video broadcasting. In Proc. ACM Multimedia'05, 32-39.
31. Wurtzel, A., Rosenbaum, J., Television production. New York: McGraw-Hill, 1995.
32. http://www.evs-global.com (2009-05-28).