Dealing with time, just in time:

sense-making and clip allocation in multi-person, multi-stream, live replay TV production

Mark Perry *(Brunel University, UK)*

Oskar Juhlin & Arvid Engström *(MobileLife @ Stockholm University)*

1. Introduction

Multi-camera television production teams face a number of major difficulties when working to broadcast live video. One of these is how TV production teams work with and across time, or at least, how they integrate historic and real-time visual content in their broadcasts. This historic material may be older information from previous games, but more often in live sports TV, it is recorded material that may have been recorded minutes or just seconds prior to its insertion into the live video feed broadcast. Teams need to do a great deal of sense-making in selecting visual feeds from multiple cameras, made all the more complex when multiple streams of recently recorded content (instant replay footage) is able to be edited and available for use within seconds of the original action of interest itself taking place. When multiple instant replay operators are all working simultaneously to generate interesting and relevant content that is competing with content from the (also potentially relevant and interesting) real-time camera streams, the problem of selection becomes all the more complex. This is the case confronting live broadcast teams in contemporary sports TV.

Why are multiple visual streams so important for sports TV producers? After all, are they not just attempting to 'show' the game? Well, certainly, they need to show what is happening, but

any but the most naïve sports viewer will recognise that sports TV is more than simply a factual account of events as they happen. Multiple cameras are deployed around stadia to show gameplay from different angles to create variations in tempo and emotional atmosphere that enliven the visual imagery, create a compelling narrative around the enfolding events, and provide an enthralling viewer experience. Yet even well-produced multicamera productions usually cannot visually explain what *is* happening in the light of what *has* happened, and instant replay is often used for this purpose, drawing from camera angles and footage that themselves may not have been shown in the broadcast. As with the multicamera broadcasts, recent developments in broadcast technology have meant that a very large quantity of content is becoming available for use and that allows multiple instant replay operators to edit this content in near real-time conditions.

In this, the focus of our interest lies in the co-ordination of live televisual broadcasts that involve significant technological support in the form of multiscreen displays, searchable hard drives, customised video editing hardware and control systems, and voice communication systems. The term *live* is used in a variety of ways when talking about broadcast, and to clarify this, we utilise the terms 'real-time' and 'live' to distinguish between their meanings and reduce any possible confusions. By 'real-time' we refer to the ongoing sequential actions as they actually occur, and by 'live', we refer to the broadcast stream as it goes out, which may include both images that are broadcast as they occur in real-time, as well as historically recorded images that are produced and broadcast alongside the real-time material. This is a subtle, but important distinction, in that what is commonly described as 'live TV' may include much content that is not real-time[1]. This chapter is therefore an attempt to understand better how and why live broadcasts

---

[1] Due to technical and legal reasons, at the receiver end, it is rare that video content can truly be described as 'live', because of the short delay introduced into the broadcast.

are made in the ways that they are, and how live and non-live video footage from multiple

participants is produced and segued together to create meaningful imagery within a coherent

narrative structure.

The chapter also draws on our previous studies on replay production that took place in a

somewhat less complex collaborative video production (Engström, Juhlin, Perry & Broth, 2010).

In this paper, we studied how a single EVS operator performed the replay production. This study

revealed how searching for a particular activity in the recorded material was conducted, during

these constrained and time-critical conditions, through collaboration in between the EVS

operator, the vision mixer, other participants in the production studio and the remote camera

team. This was achieved by the collaborative actions of the team, e.g. cameramen and producers,

who acted on and cross-referenced different temporal trajectories of the various media that they

were working on to make sense of the on-going action and to locate appropriate recorded media

for broadcast. This chapter differs from the former analysis in that it is based on fieldwork where

the EVS production is conducted by a number of people working in different teams of

collaborating operators. The focus of this paper involves much more talk, and our analysis is

therefore more oriented to the complex verbal interactions between operators than our previous

work, necessitating a stronger orientation to the concerns of conversation analysis than these

earlier studies have merited.

As a large team, replay operators are able to specialise more than a single operator on

individual camera footage, as well as attend to more focussed aspects of gameplay. As we see

here, multi-member instant replay production is based on functional separation of various tasks or

responsibilities. In practice, and as we show, the EVS editing, selection and organisation of work

in this scenario takes a very different form to that of the single EVS operator examined in our

earlier work. Importantly, the set up of replay production discussed here depends on many more

people working together, and the role of talk becomes more important. This is one of the distinguishing features that differentiate this chapter from these earlier studies, in that it is highly engaged with talk, and how this talk is used in organising visual content. The following analysis therefore provides an account and analysis of decision-making in visual media production; it is its orientation to the visual media that connects it to the other chapters in the book, but the very large size of the team involved, the number of media streams and the challenges of real-time production differentiate this kind of video work from these studies. Compared against our previous publication, this provides a worked methodological approach to further unpack both the motivations and the actions taken to produce replays, and to study more elaborate forms of the use of replays in live TV-production; it also extends these studies on the organisation of video-based interaction to focus in on the role of talk-in-interaction in this most visual of contexts.

2. Background

Our analytic framing is derived from conversation analysis (CA, Sacks, 1992), and may be best described as a form of interaction analysis (Jordan & Henderson, 1995), drawing on the CA, EM and workplace studies literatures. Following Mondada (2011: 542), we have attempted to perform a "sequential and temporal approach to understanding", focusing on the unfolding of actions and on showing how understanding is performed "as a collective achievement, publicly displayed and interactively oriented to within the production and the monitoring of action" (ibid). In this case, we are interested in turn taking organisation–not just conversational turns in talk, but, and drawing inspiration from Schegloff's term of 'talk-in-interaction', foregrounding the interaction and artefacts involved. This dual interest, in the two tethered phenomena of the

organisation of the coordination within the team, and the organization of the broadcasted images, is, we believe, a novel concern. In this case, we focus on interaction with and through the video materials that members are engaged with using. With reference to Sacks (1984), we are attempting to locate the "technology of interaction" in videowork, "to take singular sequences of interaction and tear them apart… to find rules, techniques, procedures, methods, maxims that can be used to generate orderly features in it" (1984: 413). Our use of CA is not specifically intended to identify turn taking components, but in identifying how team members' practices, organisation, settings, and technologies come together in support of their collaboration and in their production of visual content for broadcast. This has much common ground with ethnomethodology and workplace studies, and our special interest here lies in members' collaborative interactions with and through the visual content and technologies available to them as they unfold temporally and sequentially.

So what we are attempting to examine follows CA's concern with the 'organisation of interaction' (see for e.g. Schegloff, 2007), and the ways that interaction is informed by language, embodied action and the visual content that video production work involves. Schegloff (2007) provides an interesting list of features of interaction (his 'generic orders of talk-in-interaction', 2007: xiv) that we reinterpret in the light of our interest in how live and non-live video footage from multiple participants is produced and segued together. He frames these as problems without which interaction cannot proceed in an orderly way; given our topic of interest we have repurposed this list to accommodate and foreground the video streams and the visual resources used in video production, in addition to the role of talk that Schegloff identifies, and have changed his wording accordingly. We have retained Schegloff's titles in these problems to highlight common features that we draw from, although here we have framed these in the particular light of our analytic context:

(1) the "turn-taking" problem: multiple operators offer many potential video streams for broadcast. Who should be broadcast next, and when should they expect to do so? How does this affect the construction and understanding of the members' turns at talk? This is not just a problem for those making the selections, and the EVS team should not be simply seen as computational 'resources' for functionally enabling the broadcast, but as actors who have real agency in managing these turns in visual media selected for broadcast.

(2) the "action-formation" problem: how the resources of language, the body, the environment of the interaction (including visual media), and position *in* the interaction are fashioned into conformations designed to be, and to be recognizable by recipients in the production team as, particular actions–actions like requesting, inviting, granting, complaining, agreeing, telling, noticing, rejecting, and so on. That is, in the context of video production, how members' actions are artfully created to purposefully be accountable to others in their team about what they are trying to achieve.

(3) the "sequence-organizational" problem: how successive visual streams of content are formed up to be "coherent" with the visual content (i.e. a 'clip') that had been broadcast prior (or even some earlier clip/s that had been broadcast), and what the nature of that coherence is. In dealing with this problem, members must work together to organise the sequential presentation of visual material to allow viewers to make sense of what is broadcast–as it is being broadcast. For the team, the practical consequences of the sequence-organizational problem therefore involves identifying prior/next actions, and making projections and expectations possible.

(4) the "trouble" problem: how members deal with trouble in speaking, hearing and/or

understanding the talk around them, or in seeing or understanding visual content so that

the interaction does not freeze in place when trouble arises, that intersubjectivity is

maintained or restored, and that the turn and sequence of production activities can

progress to possible completion. This is a particular problem in the production studio, in

which not all of the visual content or verbal interaction is available to all of the members

because of the technical resource configurations and highly constrained physical layout

that they work within.

(5) the "word-selection" problem: how the visual and audio components that are selected as the

elements of a clip (or sequence of clips) get selected by the EVS operators as a proposal

for broadcast, and how that selection informs and shapes the understanding achieved by

the rest of the production team.

(6) the "overall structural organization" problem: how the overall composition of an occasion

of interaction (i.e. a video production) gets structured, what those structures are, and how

placement in the overall structure informs the construction and understanding of the

visual and verbal components of interaction (as turns, as sequences, etc.). In the light of

video production, this problem is one in which the members shape and make sense of the

ways that TV production is to be done in the context that it occurs within, how they fit the

micro-interactional elements that they face into the macro-scale structures of the

production system, and how they might reshape the production system in the light of

changes that occur in its doing.

It is with this set of features in mind that we begin our examination of live video replay.

We do not however, take these as a programmatic set of questions to be sequentially addressed,

but as a broad set of concerns that frame the problems that video production teams need to

practically address. In doing so, we orient to these features as heuristics, only returning to examine these features explicitly in the final discussion. In this project, we extend Macbeth's (1999) exhortation to examine the 'praxeology of seeing with a camera' from an individual perspective to that of a collective of individuals, each with their own specialised roles, in how teams do seeing with many cameras. Recognising that the TV production process depends on many more modalities than talk, our study belongs to the body of work focussing on the transmodality of interaction (Murphy, 2010), and the identification of eye gaze, gesture and body position. The attendance to the visual presentations of activities on the abundant screens, and the identification of visually recognisable member categories is of importance in the production of live TV broadcasts (Perry et al., 2009), as is the physical movement of cameras another feature of the interaction (Broth, 2004), as well as the talk mediated through the radio communication system and the body movements made visible in the tight space of the production studio.

We have no theoretical interest at this stage in how the final broadcast is watched by their remote viewers, and indeed, the corpus of data that we present here is not intended to offer insights into this. However, we have an interest in how the producers orient to the viewers and in the way the production is organised to convey a particular story, in line with recent work by Mondada (2003, 2009). Her analysis (2009) of the making of a debate, through the use of split screen broadcasts as visualization of grammatical orders in a debate and the production of ways of seeing the work, shares a common concern with ours on how the interaction resonates with the broadcast. She unpacks how the team is not just broadcasting from a studio, but how they are producing a TV debate. The editors and camera operators collaborate to construct a particular way of telling what is going on by being attentive to the talk between participants, which includes anticipation of turns at talk and the identification of potential addressees to emerging questions. Split screens are, for example, used when a question is posed to the audience, which might

include more than one person. Mondada (2003) has also shown in her studies of video production in surgical work how the producers can orient to multiple purposes that smoothly run together, such as their deployment in supporting the surgeon's need to see for himself as part of an activity, and at the same time produce a view to demonstrate what he is doing for an audience. Talk and gestures are fine-tuned to collaboratively accomplish these multiple ways of using video. In specific we show how talk supports them in "looking together" at a large amount of possible views of the same event. Talk and collaboration is used to handle and make use of an asymmetry of access to visual content, as well as providing each other insights into upcoming events.

In all, recent detailed empirical studies are beginning to unpack how interactional features, beyond just conversation, have been applied to enable the production of accounts of activities in broadcast television and video production. These studies are important influences in the continuant study of the making of live visual media, and give some useful insights into how such an analysis might be applied to our own research context. Our own contribution here is to extend this work on the production of video, in which members construct a view into the live game that they anticipate. We examine how the production attempt to meet multiple audience concerns, and how they achieve and make sense of this in a complex technical, spatial, visual and auditory setting, drawing from multiple streams of potentially broadcastable material, some of which is temporally out of sequence.

3. Setting

Data collection[2] took place during early 2010 in an outside broadcast studio (also known as a 'scanner') located in a vehicle parked a short distance away from a large, international sports stadium, in which the rugby match that forms the televised event in our analysis took place. The production vehicle can be transported quickly from site to site, and opens, concertina-fashion, to provide an enormous working area. This highly expensive custom-built mobile studio is kitted out with an extraordinary array of state-of-the-art live broadcast production and communications equipment (figures 1 and 2). The layout of this vehicle can be seen in figure 3, with our analytic focus annotated and highlighted in the two shaded areas, on the instant replay team (highlight showing camera position), and the Director (highlight showing his visual orientation). In the particular match that we present here, visual footage from 20 cameras fed into the scanner, alongside audio links, satellite and other communication feeds (figure 4). This set of visual data feeds presents a hugely complex technical exercise (see figure 5) that must be designed and tested in situ. The number of working cameras and other equipment is usually not possible to define completely due to contingencies at the site, with the consequence that all productions are likely to be unique as teams will be working with some personnel and technological configurations that they have not encountered before.

Figure 1. Production Assistant, Director and Vision Mixer (left to right)

---

[2] Permission was received from both the broadcaster and the participants to undertake this research.
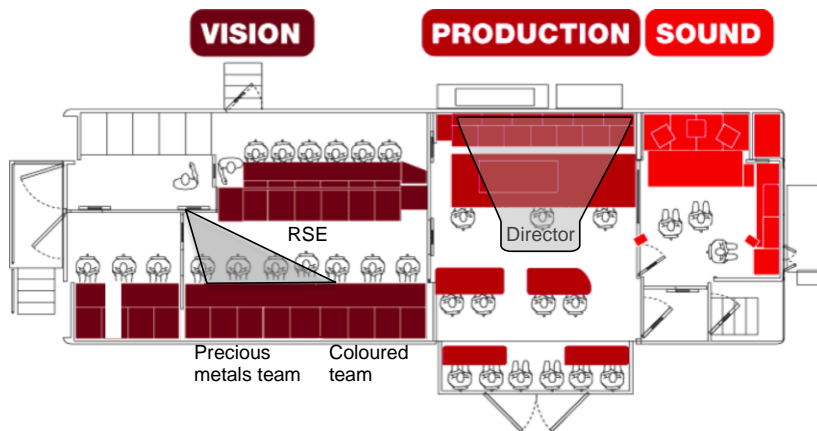
Figure 2. Gallery wall (part)

Figure 3. Configuration of the scanner (adapted from SISLive, 2011)

Two key parts of the scanner are used in image production, the 'gallery' and the instant replay suite. The gallery is composed of a wall of video monitors that display all of the visual feeds and other relevant visual information (such as time displays and match graphics), that are available for broadcast or may be used for monitoring purposes. It also consists of a desk with control panels that allow users to switch broadcast feeds or to create image transitions (wipes, fades, overlays, etc.). Several operators are situated in front of it. The most notable operator in the instance that we observed was the Director, sitting in the centre of the gallery directly in front of the displays (figures 1 and 3). The instant replay suite is a thin corridor that leads off from the production area, separated from this by a soundproof door. Here, along one wall, there is another wide bank of display monitors, lining the external wall for several metres, with seats for seven operators (figure 3). Underneath these displays sits a desk with several devices, each equipped with a number of switches, dials and lights. Although the roles of the individuals involved and their work tasks are understandably very technically complex, we will try to present a simplified overview of this below.
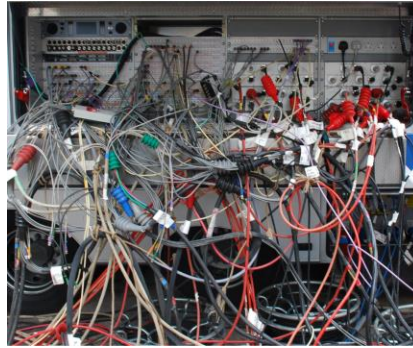
Figure 4. Outside broadcast vehicle in situ

Figure 5. associated side panel wiring

The role of the Director is to mix together various streams of available content for live broadcast. He does this by selecting one of the visual streams on the monitors in the gallery by directly pressing the button associated with that visual feed on his control panel.

Organisationally, he is able to request changes in the camerawork and ask for particular forms of replay material (where this is possible), and is assisted in his work by a production assistant (monitoring time, communicating with the commentators, etc.) sitting to his left, and a vision mixer (assisting with technical aspects of the visual presentation), to his right. He is connected by audio to the camera operators and instant replay operators (as well as other relevant production units) via an intercom.

The instant replay team is organised into two groups, who are co-ordinated by a replay subeditor (RSE) sitting at the centre of the desk (figure 3). The two groups (known as EVS operators) consist of six individuals, labelled as primary colours (operators coded as red, green, blue) and precious metals (gold, silver, bronze). Each EVS operator has a near identical set of instant replay control units (known as EVS machines) and similar configurations on their monitors (see figure 6). The replay subeditor's position is unique in that his part of the desk has a different set of equipment on it as well as a different set of display configurations (figure 7). He is

also unique, in that whilst all of the operators can hear the Director's talk from his intercom over a public address system, only the RSE can use the intercom system to speak back to the Director by pressing a button on his desk and speaking into a microphone. As will become apparent in the data that follow, several members of the EVS team recorded in the study were themselves Welsh, and keen followers of the Welsh rugby team that they were filming in a major championship match.

Figure 6. EVS team: gold, silver and bronze (left to right). Figure 7. Replay Sub-Editor (on left).

4. Analysis

The focus of our analysis lies on the interaction between members involved, showing how they organise their activities around the production of visual material for the live broadcast. These include co-present interactions between the 'precious metals' team of operators, and their interaction with the RSE, and remotely, in the audio interaction between the EVS studio and the Director, and in the Director's interactions with the camera operators. Together, the work of the instant replay operators is to select recorded footage that is relevant to the camera footage so that

replays can be meaningfully cut into real time camera footage. More specifically, the contribution of this chapter–in line with the aim of this volume–is that we show how selecting a replay segment is a local achievement, with a focus on the clip as the product of a collective action of looking and clip editing. In addition to their orientation to each others' work and actions, these actors have access to other contextual information that may impact on their situation awareness to help them make better-informed decisions about their own work.

The production team's interaction sits within an institutional context, and we must recognise that their actions and talk is framed by this institutionality, as well as being bounded by the unique spatial and technical setting that their interaction takes place within. This is not mundane interaction or conversation, in the way that much early EMCA work has examined, and this organisational context and set of professional practices are likely to have a substantial role in how utterances and actions are shaped and understood (e.g. Drew & Heritage, 1992; Heritage, 2004). This is relevant in that the forms of interaction will be shaped by the nature of the hierarchical structure of the production unit and the organisational demands of the visual work that they are undertaking, and social interaction within this context is necessarily specifically shaped by and within this particular setting through these institutional concerns. Thus, for example, the Director made the cuts for broadcast, and all of the visuals and commentary regarding the relevance of visual information therefore needed to be structured in a way that would allow him to make selection decisions. Because of the technical configuration in the scanner, he was unable to edit replays, or even to scroll through them, and could only request that these be 'rolled' by the replay operators themselves. The asymmetrical arrangement of communications between the Director to the replay team also meant that the organisational structure was embedded in the technological apparatus. Moreover, these technological arrangements were not simply a result of technological limitations, but given that they were

reconfigurable, must have reflected institutional choices in their configuration. The forms of talk

in use too were also shaped by professional practices and institutional concerns, with, for

example, the Director selecting and announcing coded camera and replay footage without

negotiation or a need for making his selections locally accountable to his team members (cf. for

e.g. Goodwin and Goodwin, 1996).

In terms of the transcription below, a simplified form of Jeffersonian transcription is used,

with speaker abbreviations for Director, commentator and bronze recorded below as Dir, Com

and Brz respectively, and the primary colour team members (off-screen, and not available for

identification) labelled as CT1, CT2, or as CT?, where they cannot be individually identified. To

accommodate for the broadcast footage, and to give some context to the verbal interaction below,

the broadcast video is added between double vertical bars with a numeric value indicating the

sequential video cuts (i.e. || n ||); the associated footage is also shown in Figures 8-22, taken on

the initiation of the cut. These images are oblique because they are taken from screenshots inside

the scanner, and the camera was not angled straight at the screen. In its totality, the single

sequence examined lasts 1 minute and 15 seconds. It is not obviously a particularly unusual data

segment, and several other instances from our dataset offer a similar set of interactional elements.

What we observe in this sequence that makes it interactionally interesting is how the production

team collaboratively watch and make sense of gameplay, and simultaneously produce topically

relevant actions and verbalisations in co-assembling image sequences. The episode is separated

into several brief excerpts, each of which is analysed in turn. These episodes are not discrete units

and are interconnected in their topical concerns, but this segmentation allows both a pragmatic

solution in providing a simple explanation to be interwoven alongside the data, as well as to

segment loosely related sets of interactions. Excerpt 1 begins with the broadcast following

camera cuts around a set of passes on the pitch; a problem in gameplay is collaboratively

identified and instant replay footage is very quickly assembled on the back of this identification:

(1) Forward pass observed (3:20)

```
            || 8 ||
01 Com:     Thomas || 9, 10 || (0.4) to Jones.
02          ((untranscribed commentator commentary continues))
03          ((loud groans/noises from EVS team))
04 Brz:     || 11 || ohhh, he's forward ((raises left hand outstretched twd her screen))
05 ?RSE:    blerrr, ... ((refs whistle blows))
06 Dir:     couple of turns
07 Brz:     f|| 12 ||orward.
08 Brz:     forward [pass. yeah. on me
09 Com:            [forward pass
10 Brz:     ON BRO::NZE, (.) ON BRO::NZE
```


Figures 8-12 (read left to right, top to bottom)

What the footage shows in this sequence is the development of a tussle for the ball (figure

9) that results in the ball being passed to a red player. The ball is passed on to another player who

catches it and makes a dash for the 'blue' end of the pitch (figures 10, 11), culminating in an

apparently scoring 'touchdown' (figure 12). Yet the EVS operators' and subsequently the

commentator's talk around these images demonstrates that all is not as it would visually appear.

After the commentator (Com) begins with a description of passing between players (01, 02) and

on-going play, some of the team members' begin to make audibly recognisable sounds of despair,

mimicking, as it were, their broadcast audience. This is followed by Bronze (04), who notes a

particular comment on the game: 'ohhh, he's forward', again followed by someone saying "Blerrr!" in what approximates to a sound of disappointment just prior to the whistle being blown. These are clearly comments on game play, and given the way that they are temporally aligned to poor Welsh play (most of the EVS team being Welsh nationals), are partisan.

All of the broadcast footage here is live camerawork (figures 8-12), showing different angles and levels of zoom, yet these camera angles are poorly framed to show whether the pass is forward as it is hard to ascertain the action in relation to the pitch (see figures 9-12 above). However, at the same time as this, the EVS operators are also watching the cameras that they are responsible for in real time and are able to rely on their insider knowledge of the orientation of these cameras in relation to the pitch. Here, what we can unmistakeably see in the talk at this time is that the team are making sense of what is happening on the field, and drawing the attention to others of their own observations as they develop a stronger sense of certainty about them. This can be observed in the initial 'forward pass' comments (07-09), followed immediately by confirmation of this as the EVS operators begin to review their recorded footage (notably by Bronze beginning to replay her footage immediately before 07). At this point, our analysis has now established the form of professional vision held by the team, illustrating how they do *looking together* in making relevant clip selections, but also identifies their orientation to an audience-like vision, in looking with some of the same concerns as their TV viewers.

A forward pass is an important and illegal feature in the game, and is important to the production team because of its consequences for filming gameplay. However, it is important for them not just because it will result in a game stoppage if this is recognised by the referees. It also means that other things will need to be brought into play in the production. They need to recognise that the referee may (or may not) see this foul, and that this will need to be returned to for analysis in a moment of stoppage (which will invariably require instant replay footage).

Indeed, because they recognised this before the referee blew his whistle and stopped play, the

EVS team were able to make predictions of the next possible turns that the game might take.

More than the use of this knowledge later on in the game, there are some immediate

consequences that they have to attend to here. Gameplay-relevant graphics may need to be

overlaid on the live broadcast for the audience. In case of a longer stoppage, footage needs to be

available to fill in for time in which little interesting happens, and to add narrative to explain the

reasons for stoppage. But there is another reason for the calls made by the EVS operators and

team beyond simple reuse of footage, and this is in allowing the Director insight into what is

happening *ahead* of the foul being called by the referee. This allows the Director and RSE to

begin to set up a sequence of actions in advance of their being required, or at least as early as

possible. Attending to the vast number of screens in the gallery composed of the 20 cameras

available, the six EVS streams, graphics, and telestration capabilities (see figure 1) and directing

their work to interesting opportunities obviously requires a huge and extremely demanding feat

of mental agility, even for an experienced professional. Knowing ahead of time what to attend to

and what to ask the various operators of the component parts of the system to do has an obvious

advantage, and as we see, this produces a situated practice that is enacted by the other members

of the team in searching for what has preceded the foul. So by announcing the foul ahead of its

call by the referee, the members are acting to reduce the number of options that the Director and

RSE have to attend to. In effect, this allows them to plan ahead in what is a largely uncertain and

unpredictable terrain. The sequence then continues, with the analysis showing how this

information is made use of in subsequent verbal interactions and image selections:

(2) Making proposals for broadcast (3:32)

```
11 Dir:    tel[estrate,
```

```
12 ?RSE:      ['nything else?
13 Dir:    TELESTRATE
14 ?:      err
15 ?RSE:   gold
16 ?Gold: ye- yeah gold
17 Dir:    gold
18         (0.5) || 13 || (0.5)
19 ?RSE:   gold (.) then bronze (.) || 14 ||
```


Figure 13   Figure 14

What begins in the live broadcast footage with a close-up of a player (figure 12) following

a disallowed point score (called a 'try' in rugby), is followed here with a cut to the linesman

(figure 13) who identified the forward pass, and cut back to the disallowed try scorer (figure 14).

These could be considered as 'normal' patterns of video cutting, as they demonstrate first the

rationale for the referee's decision, followed by a close-up on the disappointed player. That these

sequences are normally anticipated means that the EVS team could be reasonably expected to

have some time before their replays are due to finalise their edits and to simultaneously announce

their readiness to show pertinent replay material. So at the start of this second excerpt, the EVS

team are prepared for an extended period of stopped play and the Bronze EVS operator has a

prepared replay segment ready to roll for the Director. This has now been announced multiple

times, both as Bronze spots what she thinks is a forward pass (04, 07), gains confirmation from

her team (09), and this is announced loudly over the intercom to the Director (19). Bronze EVS

has also announced that it is visible in her own footage ("on me", 08), and having the only

woman's voice this is easily recognisable as her, and she reinforces this by loudly announcing

"on Bronze. On Bronze" (10, from excerpt 1). Notably over this period, the fact that Bronze has

picked up the replay footage is not recognised explicitly by the Director. However, he does now

make a request that shows he has recognised something has occurred in the gameplay, as he calls

out over the intercom: "Telestrate" (cueing this on 11, then announcing the cut to telestration in

13). This is not a call for Bronze though: telestration requires instant replay footage from an

overview camera (such as Gold) so that the commentator can annotate it using a stylus and this is

broadcast in real time (i.e. televisual-illustration). Thus we can see that he recognises an

important event, and that it is likely that a reasonably long game pause will occur in which this

telestration can be slipped into the broadcast. Such a call could be reasonably expected by

Bronze: it would be unusual to cut directly to the close-up action. Instead, an overview replay

would be shown in which the commentators could show that, how, and when the foul occurred

(the forward pass). Only then would this be followed up with a close up shot from the side (by

Bronze), from which it would be hard to determine who was in front of whom were it to be

shown on its own without prior overview footage to add this context. This kind of 'topic

orientation' is common, and we have observed it in previous studies of ice hockey TV

productions (Engström et al., 2010; Perry et al., 2009). This discussion continues in excerpt 3, in

which we show how this observing, clip production and proposal, is followed by new visual

content from the live cameras and replays that offer to displace Bronze's apparently increasingly

less relevant replay:

(3) Running Gold and 'bumping' Bronze (3:38)

```
20 RSE:     go on. run that (.) I can't see it (.) run
21 ?:       gold
22 Dir:     stand by vt gold
23 ?Gold:   gold
24 ?RSE:    || 15 || Stand by vt gold
25 Dir:     run gold
26 RSE:     run it || 16, 17 ||
27 ?CT1:    reaction on green
28 RSE:     one on bronze as well. one on bronze
29 CT1:     another good reaction! ((off-camera))
30 CT2:     yeah ((off-[camera))
31 multiple voices:   [reaction (.) REACTION on gree:n, ((off-camera))
```

```
32 Dir:    °absolutely°
```


Figure 15   Figure 16   Figure 17

       Here we see Gold's replay cued up and selected, as expected. This is the first replay we have seen broadcast so far. This is preceded by a 'swipe' (a static image is briefly broadcast, see figure 16), used as a common indicator to audiences that the broadcast is no longer 'live'. What we then observe is that during Gold's selection, a new and unexpected turn starts to arise with a call for Green to be shown for its suitable display of visual 'reaction' by the players (27). It provides a detailed shot in which it is possible to see the facial expressions of a player. This appears to displace the 'promised' ("Gold then Bronze") upcoming use of Bronze's replay material, and Bronze produces an immediate verbal response ("one on Bronze as well. One on Bronze"). As we can see, what is announced is not always what is implemented, showing that the sequentially of image selection is not deterministically tied to talk, and that these offerings are not so much promises, but are more advisory in nature. Bronze evidently recognises the clip competition from Green, as we can observe from her demarcated "as well" (28) that is in direct reference to the alternative clip. The close sequential placement of this turn is clearly produced as a response to the new and competing proposal for showing Green–given that they are competing for limited broadcast time, the likelihood of using both replays is low. What follows is something of a verbal scramble for attention from those representing Green's replay offering. Several loud voices call for Green's footage to be used (31). However, it is unclear to everyone how the Director reacts to this: we are presented with a quietly intonated response ("absolutely", 32) that appears to pair with this call for Green's selection, yet he does not select another replay,

remaining on the still rolling overview replay from Gold (still seen in figure 17). This excerpt

makes visible the importance of verbally negotiating for individual camera selections. It is

important here is to understand here that there is an asymmetry of access to visual content

between the production team members. The Director can only see the first frame of their

selections in his gallery, and not the full replay sequences: he cannot know what the replay is, or

what it will later show, just from what he can see on his monitors in the gallery. Thus we see a

distribution of labour as multiple persons do the work of selecting cameras for EVS, and this

comes to depend more on the ways in which the replay operators can argue for their case than on

the static and uncertain visual information available to the Director. The negotiation then

continues:

(4) A call for replay 'reaction' (3:53)

```
33 RSE:    another one on bronze
34 Dir:    going back to 3 (.5) || 16 ||
35 Dir:    w||18||e'll do another one, give me ||19|| another one, number 2
           ((refers to camera 2))
36 RSE:    on bronze
37 Dir:    ok stand (0.5) and reaction
           ((cameras directed to spectators))
38 RSE:    hold it there
39         ((long pause))
40 Dir:    so come back for the penalty
41         ((long pause))
```

Although the RSE makes a call for Bronze to be selected (33), the Director gives a

forewarning that camera 3 will soon be selected (34), immediately after this. Here, following the

de-selection of Gold's replay (32) to a live camera, only live footage is used in this part of the

broadcast, shifting from cameras 3 to 2. In the EVS studio, the replay search continues, although

as can be seen in the audio transcript, nothing interesting appears to be found. This can be

observed as the EVS operators keep scanning through replay material without stopping to pick

out any selections or verbally commenting on their search, although the RSE vocalises (also

available over the intercom) that Bronze is still ready to present content (36). The Director then

makes a request to the EVS team to propose another reaction shot (37); notably, he does not

select an existing replay that has been proposed, but makes visible a division of labour in the

production team. Through this request, the EVS team know that they will be able to select a

single 'reaction' replay clip based on a judgement of the content available to them and propose it

to the Director (most likely by RSE via a verbal proposal over the intercom). The RSE then

notices something interesting in the scrolling footage of one of the EVS operators (38), and (as

we later see more obviously in the excerpts that follow) ties his conversation with one of the

coloured team's activities on his screens, matching his talk to the timing of the images playing

out on the screens in front of them:

(5) Bronze is selected (4:07)

```
42 ?:     ((ref is seen to play penalty card in real-time footage)) OEU::AGH!!
43 Dir:   y:ep
44 RSE:   show that again (0.6) that angle (0.4) run it from there (.)
          run it from there
45 Dir:   [come on in 5 ((calling camera 5 to zoom in)) || 20 ||
46 RSE:   [stop stop stop
47 Dir:   what is it? (0.4) bronze?
48 Brz:   yeahh
49 RSE:   go back (0.7) go back
50 Dir:   stand by bronze? (0.4) and run bronze || 16, 21 ||
```


Figure18   Figure 19   Figure 20

This section begins as one of the players is penalised and this is followed by what appears

to be a groan of disappointment (42). This is known to be a topic of concern for the production

team, as this is anticipated as being topically relevant to their viewers. Penalties typically result in

a discussion by the commentators, and may (depending on the length of time available before play resumes) offer an opportunity for replays of the event leading up to the penalty awarded to examine the visual details of the contested event on the pitch. The RSE, who had been commenting on one of the coloured team's replay edits, now takes a much more active role, requesting to see the footage again, and suggesting edit points ("run it from there", in 44). As he says this, his speech is co-ordinated with the rewinding actions of coloured team EVS material (which can be seen in a 'tile' on his own display), so that the coloured team's material onscreen actions are temporally tied to his talk; this continues in 46, as he calls on them to "stop stop stop", and again at 49 (and in excerpt 6, in 52). Bronze then finally gets to play her material (50). This is an astonishingly long period to have retained replay footage onscreen given the normally fast pace of search and video editing activities seen in the room up until this sequence, and indeed by the other EVS operators during the time: her screen remains with the same edited image static onscreen from 15.5 seconds into our video sequence until exactly 1 minute in (clip timing at 4:40). She does this without any direct instruction to keep this active, and we can only surmise that this is done because, a) the repeated calls for "Bronze" by her, the RSE and Director are not completed with its subsequent broadcast, nor though are they dismissed as they are continuously referred to by members, and b) that this replay is continuously assessed and deemed to be useful and highly relevant to the unfolding game, despite the time elapsed and other game-related activities that follow on.

The excerpt also shows how the actual work of looking at the material and deciding on the editing becomes a collaborative effort. The RSE then gets additional visual resources to suggest what should come next. However, such re-editing of the EVS operators' selections takes additional time, which, in this case, was available. In the final excerpt, we see a clear reference to the collaborative nature of attending to temporal concerns by the RSE, as well as to the

relevancies of the game dynamic of fitting the replay into the on-going broadcast without

disrupting the viewers' experience of real-time action as the rugby game plays on:

(6) Planning replays for a possible future outcome (4:20)

```
51 ?CT?:   and its reaction's on silver Gav ((man's name))
52 RSE:    go back again
53 ?CT?:   reaction on green
54 RSE:    we'll telestrate that for next time, if, if there's a
           stoppp: || 22 || :age, OK?
55         ((with the cut to fig. 22, the broadcast fades into a second replay))
56 RSE:    just keep that up there. OK?
57 RSE:    OK
58 RSE:    if they score (.) if they score we'll do it
           ((broadcast cuts to live camera))
```
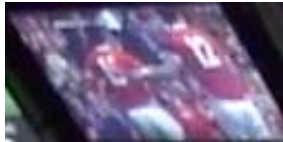
Figures 21   Figure 22

        Here, we can observe another feature of broadcast in the conversation, as the team do

work to find replay footage that fits to the time available for replay. We see this less explicitly in

the fact that Bronze failed to get selected earlier for a long time despite being relevant, ready and

waiting, but the RSE's reference to a stoppage (54), and then more precisely, to a goal (58), is a

clear verbalisation that new footage is to be kept available to be fitted in. They explicitly do this

when a) there is a natural break of play of a suitable length available to insert pre-recorded

content, and b) that it is relevant for the game narrative, returning to recorded footage as a

resource to explain the role of past events on present ones (i.e. as we see in 58, "if they score").

        The data sequence examined, we now turn to pull out aspects of the interaction at a more

abstract level.

5. Discussion and conclusion

The analytic focus of this chapter lies in examining the alignment of visual content and talk in organising members' interaction around video clip allocation to generate a coherent image-based narrative, and showing how they make themselves accountable for their selection of video clip proposals for broadcast. It is worth returning to Schegloff's (2007) 'problems' at this point, as these pose the core issues that the production team face. Taking these in turn, we have shown, in the instance of this perspicuous case, how the participants achieve this.

From the mass of real-time camera footage and recorded content–only a part of which is visible to all of the members involved–the team manage to agree, in a very rapid way whose content will be broadcast in visual "turn-taking" (while recognising that verbal turn taking is important, we focus here on the video material). It appears that this is tightly tied to the "sequence-organizational" aspects of action, as successive visual streams of content are made coherent. The participants do so by orienting to some apparently simple interactional features. Narrative features of the game dominate here: certain image selections are visibly understood as normal practice, most likely because they are commonly recognised as those actions that will make sense to the viewer. Showing activities that have sequential relevance, or typical patterns (e.g. cutting from penalty to referee, to players involved, to replay long shot, to replay close-up) clearly fall into this category, so that the Director and EVS operators both have an idea of what types of footage is likely to be called for. Here, we also see the "word-selection" problem addressed. As new visual elements are layered into the live broadcast, these delimit the sequential relevance of items that might reasonably follow them, as does the expected time available for them to be played within (e.g. "we'll telestrate that for next time, if, if there's a stoppp::age", 54). Judgements are being made about the reasonableness of potential visual turns (i.e. visual content

> **Kommentar [XX1]:** Ok, so it wasn't a mistake. Stop:: age?
> (MP) fixed

made available for broadcast) in the context of the on-going broadcast footage and their relevance

to the game (as can be seen in Bronze's continued holding of her footage).

Nevertheless, other non-narrative concerns also guide these orientations to content for

selection. There is an expectation that footage has to be available for use 'in time', but this is not

always practically possible and the broadcast needs 'fill in' footage until the relevant replay

material is ready. Similarly, multiple camera angles from different EVS operators may be

available and decisions need to be made on which replay operator to select from. When the

Director cannot attend to these directly (as in this case), the RSE makes selection suggestions,

and, as we show in the data, this is handled conversationally in combination with the visual

resources by the EVS operators and RSE. To do so, EVS operators have to identify and

communicate that they have a relevant sequence (and what it is) very quickly so that it can be

fitted into the real-time camera footage. The Director verbalises what he is about to try to do.

This allows the replay team to know who is likely to be selected next and gives them an

indication of whether to edit or sit on existing footage, or to move onto new material. Here, we

begin to see the "action-formation" problem being addressed: shot proposals, requests,

agreements, and noticings being presented as recipient-designed recognisable actions. In the case

of the Director, these interactions largely occur through the intercom, and are necessarily verbal,

but in the EVS team, these are more indexical, and talk is clearly tied to the visual imagery

present or not present. As we can see, because of this collaboration, the team can review more

video material by functionally separating their activities than if they all have access to the same

visual resources. However, this division of labour and subsequent selection of an individual

operator's material depends on what they are able to make sense of through their talk. This talk is

not always sufficient for this purpose, which makes them revisit and re-edit the visual material in

a reflexive cycle of live video watching and recorded video editing.

Kommentar [XX2]: Ok?
(MP) yes

A great deal of organisational work has been deployed to ensure that "trouble" in the interaction does not occur. The organisation of the material resources (intercom, gallery, EVS screens and their layout, RSE screens, scanner floorplan), the team's clear divisions of labour and their pre-match planning are all designed to ensure that problems are minimised. The data present few obvious or terminal troubles that interfere with the production, perhaps as testimony to the skilled professional practices and configured material environment that they work within. As we see, the "overall structural organization" of the unit is rather different to the emergent structure of talk and interaction in non-institutional settings, and this environment provides a setting that constrains and scaffolds patterns of communication, and delimits possible interpretations that can be layered onto their interaction. We see this, in a very simple example, in the Director's call to the EVS operator to "run gold" (25). There are no conversational niceties present in this talk, no narrative explanation to make this action accountable or addressing the immediately previous talk, no explanation of what "run" means or when it should be enacted, or of its closure in an offering of thanks for its presentation. The Directors' words are instructive and do not open an opportunity to extend more talk in a following turn. This talk is brief, utilitarian and formed to meet the specific, situated demands of the task-at-hand, as well as reflecting their 'differential access to resources and power' (Heritage, 2004: 114). During this kind of talk, members are orienting to their institutional identities, although we also see some slippage between their institutional and ordinary identities (e.g. 32, 42; see also Heritage, 2004).

So, in conclusion, the analysis presented here allows a number of features of the setting to be drawn out and examined. We show how team members share video selections, both in co-presence and remotely, so that their image selections are made observable and negotiable. We also show how they compare multiple video streams that appear to have timely relevance, and examine how they interrogate these competing selections to determine which, in which order, and

Kommentar [XX3]: remind us of an example of this here, so as not to remain in the abstract?
(MP) fixed

when they are suitable for segueing into the live broadcast. Such decisions on image selections

may be accounted for with reference to their apparent intelligibility, narrative relevance,

timeliness and temporal length, as well as their aesthetic qualities, although these factors are *not*

always visible to those making the selection decisions. This may be because either too many

parallel video streams are visible for the RSE to attend to, and the fact that they are too long to

fully review in the time available, or in the case of the Director, because these streams are simply

not visually available. The members arguing for their own selections are therefore accountable

for the relevance of their footage (after the event), and as we have seen, multiple members may

also call for selections to be made, making the proposal appear more warranted. This analysis of

the work of instant replay shows how the members of a large and distributed team are able to

produce meaningful and topically relevant material from a number of real-time camera streams

that is suitable for interpolating into a live television broadcast. To achieve this, the team must

attend to multiple streams of both live and recently recorded visual content, and to use talk and

their produced visual resources to make sense of and coordinate their actions. A close analysis of

their work shows how this is achieved, referencing an adapted version of Schegloff's generic

orders of talk-in-interaction that integrates the visual and verbal resources that they have to hand.

References

Broth, M. (2004). The Production of a live TV-interview through mediated interaction. In *Proc.*

   Int. Conference on Logic and Methodology. SISWO, Amsterdam.

Drew, P., & Heritage, J. (1992). Analyzing Talk at Work: An Introduction. In Drew and Heritage

   (Eds.), *Talk at Work* (pp. 3-65.). Cambridge: Cambridge University Press.

Engström, A., Esbjörnsson, M., Juhlin, O., & Perry, M. (2008). Producing Collaborative Video:

Developing an Interactive User Experience for Mobile TV. In *Proceedings of UXTV

2008: First International Conference on Designing Interactive User Experiences for TV

and Video*, Oct 22-24, Mountain View, CA, USA.

Engström, A., Juhlin, O., Perry, M., & Broth, M. (2010). Temporal hybridity: Mixing live video

footage with instant replay in real time. In *Proceedings of ACM CHI 2010*, Atlanta, GA.,

USA, 1495-1504.

Goodwin, C. and Goodwin, M.H. (1996). Seeing as a Situated Activity: Formulating Planes. In

Engeström and Middleton (Eds) Cognition and Communication at Work (pp. 61-95).

Cambridge: Cambridge University Press.

Heritage, J. (2004). Conversation analysis and institutional talk. In Sanders and Fitch (Eds.),

Handbook of Language and Social Interaction (pp. 103-146.). Mahwah NJ, Erlbaum

Jordan, B., & Henderson, A. (1995). Interaction Analysis: Foundations and Practice. *The Journal

of Learning Sciences*, *4*(1), 39-103.

Lee, J.R.E., & Watson, D.R. (1993). *Interaction in urban public spaces* (final report).

Manchester: Department of Sociology, University of Manchester.

Macbeth, D. (1999). Glances, trances, and their relevance for a visual sociology. In Jalbert (Ed.)

Media studies: ethnomethodological approaches (pp. 135-170). Lanham: UPA.

Mondada, L. (2003). Working with video: How surgeons produce video records of their actions,

Visual studies, *18*(1), 58-73.

Mondada, L. (2009). Video recording Practices and the Reflexive Constitution of the

Interactional Order: Some Systematic Uses of the Split Screen Technique. In *Human

Studies*, *32*, 67-99.

Mondada, L. (2011). Understanding as an embodied, situated and sequential achievement in
interaction. *Journal of Pragmatics*, *43*, 542–552

Murphy, K.M. (2010). Shaping Things to Come: Temporality and Transmodality in Design,
University of California, Irvine, Paper presented at ICCA10, Mannheim, July 4-8, 2010.

Perry, M., Juhlin, O., Esbjörnsson, M., & Engström, A. (2009). Lean collaboration through video
gestures: co-ordinating the production of live televised sport. In *Proceedings of ACM CHI*
2009, Boston, Mass., USA, April 04-09, 2279-2288.

Sacks, H. (1992). *Lectures on Conversation*, Volumes I and II. Edited by G. Jefferson with
Introduction by E.A. Schegloff. Oxford: Blackwell,.

Schegloff, E. A. (2007). *Sequence organization in interaction: A primer in conversation analysis*,
Vol. 1. Cambridge, England: CUP.

SISLive (2011). OB 3 HD Production Units – Outside Broadcast. http://www.sislive.tv/outside-
broadcasts-ob3.php. Retrieved on 02/7/2013.