Video Composer and Live Video Conductor: Future Professions for the Interactive Digital Broadcasting Industry

Richard Wages [1], Carmen Mac Williams [2], Stefan M. Grünvogel [1], Georg Trogemann [2]

[1] University of Applied Sciences Cologne, Betzdorfer Str. 2, 50679 Cologne, Germany
[2] Academy of Media Arts Cologne, Peter-Welter-Platz 2, 50676 Cologne, Germany
{wages, macwilliams, gruenvogel}@nomadslab.org, trogemann@khm.de

Abstract

Innovations in hardware and network technologies lead to an exploding number of non-interrelated parallel media streams. Per se this does not mean any additional value for consumers. Broadcasting and advertisement industries have not yet found new formats to reach the individual user with their content.

In this work we propose and describe a novel digital broadcasting framework, which allows for the live staging of (mass) media events and improved consumer personalisation. In addition new professions for future TV production workflows which will emerge are described, namely the 'video composer' and the 'live video conductor'.

1. Introduction

Television is undergoing a historical change. Interactive Digital Broadcasting will be reality in 2010+. Video material will be generated by TV producers in abundance, which will overwhelm both the broadcasters as well as the consumers. Current TV formats and forms of broadcasting do not satisfy the personal moods and interests of the consumer.

Television originated from theatre and film. Both media (and art) forms tend to be of a primarily entertaining nature and entail a *lean-back* mentality of the consumer [Melinger]. To expect that consumers will simply switch to a complete *lean-forward* mentality – like they do when sitting in front of a PC – once iTV becomes reality is an illusion [Freeman]. Moreover it is not desirably either, since both, producer and end user, are overwhelmed by the mere quantity of live video material provided.

On the other hand art has a long tradition in dealing with the question of variability in the process of presentation. Already at the end of the 19th century random processes were introduced as a deliberate aesthetic element. Early experiments on the basis of electronic machines were carried out in music in the middle of the 20th century. Young Artists tried to overcome the extreme determinism of serial compositions. Under the term *aleatoric composition* musicians like Karlheinz Stockhausen, John Cage, Pierre Boulez and Earle Brown worked on more open methods for composition. Openness in art means that certain details of a piece are left undefined by the composer. The conductor (i.e. the interpreter) is left with the task to fill the gaps while performing the piece. Another possibility is to determine the undefined parts by using the outcome of random processes. But randomness in art does not mean that the resulting work is completely unforeseeable. Art uses random processes as a medium to achieve variable but still planned and well-formed structures. Randomness is used as a generative method to achieve structural freedom and to attack the stiffness of closed works.

Bearing in mind the existing (and future) infrastructures of digital video broadcasting when bringing this kind of variability to the medium TV we hence propose the development of a TV environment, which allows for the establishment of 'virtual personalised channels'. To do so, (live) semantic annotation of video material as well as methods for live staging of media events have to be designed. The result is a drastically different – more evolutionary – process of content production and a different form of consuming. This evolutionary character will give rise to multiple variations and categories of possible broadcasting and presentation formats, which are better suited for satisfaction of individual human needs. The approaches outlined in this paper are the basis for our upcoming IST research project LIVE.

2. Related Work

Due to the growing technical possibilities the today broadcasting industry is able to offer a huge number of different channels. Nevertheless we can observe two main tendencies due to the '24 hour per day restriction' for every channel:

- Channels serving and restricting themselves to a singular and very special category, e.g. Golf, news loops, cartoons, region etc. (i.e. 'Greatest Common Divider' approach)
- Channels serving a mean set of interests trying to reach the mass audience (i.e. 'Least Common Denominator' approach)

Of course there are some exceptions to these tendencies. Broadcasters like some public stations try to combine broad overview with intermittent in depth coverage (e.g. arte theme night). But the current setting of autonomous media streams (i.e. 'channels') will not lead to a satisfaction of the individual consumer's needs. And obviously the number of channels serving certain moods are (and will always be) too small compared to billions of individual moods. A 'solution', to offer one channel per single consumer, again is an illusory vision (too many consumers, too few 'classical producers'). In addition each single end user has varying (and unknown) moods. To offer 'video on demand' (mainly feature films) is just a fragmentary answer to this problem since the hunger of being part of the world community and the coverage of live events cannot be satisfied. Hence solutions are urgently needed for the treatment of such live events for future digital broadcasting.

For the creation of iTV productions some commercial products (authoring tools) are available. One of the products, 'Cardinal Studio' from Cardinal Information Systems Ltd. [Cardinal], allows for the design of MHP-based iTV applications. Possible applications which can be created with Cardinal Studio are enhanced digital teletext with embedded graphics, news-tickers displaying the latest news and stock headlines, weather service with a world weather overview, electronic program guides and simple trivia games. Services like this are provided by almost every public broadcaster who is capable of transmitting iTV productions. With these applications the limits are already reached, since the application only allows for the editing of the graphical layer on top of the audio-visual content.

In order to realise a TV application, which accompanies the content, the described system is not sufficient, because the interrelationships of the TV content cannot be synchronised for the final application. For such cases 'on-Q Create Suite' by Ensequence [Ensequence] offers a solution. It provides a timeline to arrange the content. It offers a huge range of solutions from organising the TV program to defining the behaviour of interactive elements such as buttons and texts and assists the author during the entire workflow. A comparable product to realise the above described scenario is 'Modelstream' by emuse-technologies [Emuse].

All of the above products do not support more than one TV program and therefore do not provide any way to create a relationship between parallel TV content streams. Their main focus is the graphical presentation and user interface. Since they all were developed for linear TV productions they do not consider the new possibilities of nonlinear productions or events with parallel streams.

One approach to link the different audio-visual contents is done in the DVD production 'Switching' by Oncotype [Oncotype]. All through the story the actors pause for a moment in order to give notice to the viewer of the possibility to switch to another 'story branch'. Thus every user is following her or his own interest, creates an own path through the story and defines the duration of the story. Nevertheless this intriguing approach is not suitable for television because of the time-based nature of broadcasting. For television the story of every viewer may be variable in its path but not in its duration.

In [Molander] the scheduling and automation system of the BBC's interactive TV playout system is presented. iTV events can be activated by an operator, by a calendar or by filters listening to incoming data from the automation systems. This rule based system is fully integrated in the broadcasting chain.

For combining media streams in a meaningful way a description of the media content is required. To do so media content has to be described in a multi-perspective way, ranging from description of camera angles for a single shot up to the semantics of a movie or the action of humans within a video. In Marc Davis iconic visual language *Media Streams* [Davis] this issue was addressed for offline annotation and repurposing of audio and video content. It is an open question if this approach is suitable for TV productions, and in particular if this approach can be use for annotating live events in (near-)real-time.

A promising approach for the personalisation of broadcasting content is described in [Rhambia]. With the help of a *trellis graph* (i.e. a directed acyclic graph arranged on a timeline) a storyline is modelled. Nodes represent content segments and edges between the nodes represent permitted transitions between these content segments. This formal model is integrated into an overall framework, where adaptive content is produced at the broadcaster side. Yet it is still unclear, how this approach can be adapted for the production of live

events, here exact knowledge about the time when certain media content is produced is not available and also the semantics of the events can not be anticipated accurately.

3. Staging of Live Events for Future iTV

A more evolutionary character of the envisioned future digital broadcasting environment entails three key elements:

1. Variation: The existence and availability of a range of video material possibly dealing with one and the same event.

2. Selection: Out of this range consumers and professional users (live video conductors) choose only certain individual video objects *during the running broadcasting scenario*. It is important to stress this latter point, since it is this very type of live selection, which gives the environment a performative character in contrast to usual, pre-production and pre-selection authoring methods in the TV industry.

3. Recombination: 'Successful' (in terms of extensive use) video objects are reused, cited and rearranged during future staging processes. Due to the huge variety of user and consumer interests most of the video objects will 'survive' and only a few might get lost forever (in analogy to natural evolution). It is important however, that the environment 'is able to utilise' these objects, meaning, there has to be an abstract description of the video object (the actual representation).

3.1 Staging and Virtual Personalised Channels

Putting the consumer in the centre of our research we ask ourselves how we can create an environment which allows for personalised TV in covering mass media events. Our approach to this dilemma is to offer 'everything' (i.e. all, possibly unedited camera streams) and at the same time provide the prerequisites for the establishment of 'virtual personalised channels'. To do so we have to realise the following:

1. Production methods and tools for 'staging' or 'conducting' a live event have to be conceived. Design principles for staging live media events and for content creation have to be developed and workflows have to be defined which support the collaborative staging by professional users. The interfaces have to hide the underlying complexity

of the media objects and the workflow requirements.

2. Personal user interests have to be linked semantically with video content. Within a robust framework, which allows for the live staging of media events, methodologies for staging and content research, content oriented detection, extraction and annotation of video material and the personalisation of the users [Lee] have to work together and enforce each other in an intelligent way. The main effort will be to develop an open framework for broadcasting environments, which utilises the interaction of cognition based content knowledge with social based consumer knowledge (connect item based knowledge with user based knowledge). Scientific questions to be answered will be:

- How/what will be annotated?
- How do we link the annotated video objects (consumer/ video conductor)

The goals of this research theme are to develop new methods, tools and interfaces for detection, extraction (of knowledge) and annotation of video material and other resources. A special focus is the 'in-vivo generation of knowledge', i.e. annotation metadata is created while content is created. More specifically this will include

- Automatic offline and online detection and extraction of knowledge from media archives and from live produced media and the annotation of media objects
- Semi-automatic online annotation of live video material. Human annotation with meta-knowledge about content will be used for constellations where automation is not possible (e.g. social meaning).
- Automatic integration of knowledge and content from external resources
- Integration of all the above methods to support the staging process

3.2 Future TV Work Flow

Equipping the audio-visual component of an intelligent video object with an additional component which holds (invisible) semantic metadata radically changes the process dynamics of producing, broadcasting and consuming. Now all essential prerequisites are given for a more 'evolutionary' character of the production process:

Instead of broadcasting just a single, carefully edited video stream chosen by a producer, several (or even all) camera streams are broadcasted.
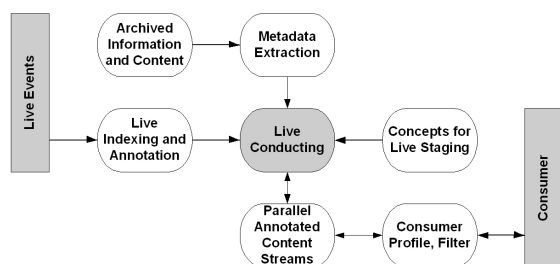

Figure 1: Live Staging Environment Overview

Besides the mere increase of the number of 'allowed' images this means the availability of an entire set of images for a particular incident. The variability of these (possibly imperfect) images allows for a direct or indirect (e.g. mood based) selection and a recombination not only by the professional producer, but also by the individual consumer.

As a consequence new and 'personalised' perspectives (or formats) evolve as an additional value for consumers. Closing the system by a feedback of consumer desires (e.g. annotation information) to the on-site producers (or even camera men) allows for a replication loop through the consumer's influence on the generation of new images.

3.3 New Professions for the Digital Broadcasting Industry: Video Composer & Video Conductor

The vision developed above gives rise to the emergence of completely novel professions such as *Video Composer* and *Video Conductor* as an analogue to a music composer and conductor who stages live events for an audience. Staging for digital TV is comparable to composing and conducting a master piece of music for an orchestra. The music composer as well as the conductor have the parameter of time and the possibility of using several different instruments playing parallel at the same time to create a master piece of music. The composer writes the piece and the conductor directs the live music event.

During staging for digital TV the video composer and the conductor are also restricted by the time limit of the TV broadcast and they can use parallel videos to create a nonlinear story composition or a live media event. The video composer and the conductor can synchronise and link the parallel video streams at special points in time to allow the

viewer to switch channels due to her/his interests. The added value for the user is a deeper emotional and intellectual experience by intuitive mood based navigation and the serving of her/his individual interests.

A remaining key question for the switching concept is: how can the video composer and conductor stimulate the viewer to actively switch the video stream at the right time without disturbing her/his immersion into the story or the live media event?

4. Design Principles for Staging Live Events

The role of the video composer is to investigate and develop the different design principles for staging live media events in the off-line mode whereas the role of the video conductor is to test out these developed methods of staging during the live broadcast (and maybe even invent new ones by improvising during the live show).

Staging live media events means for the video conductor to create a nonlinear multi-stream video show in real-time, which changes due to the interests of the end user (consumer). Today, with the introduction of digital TV it is easy for public broadcasters such as BBC, ARD and ORF to offer multi-stream videos (digital bouquet) about mass media events such as the Olympic Games, because live and archived material is existing in abundance. Apart from the time-driven nature of TV (in general 'rewinding' is not possible), it is difficult for a consumer to find the info she/he is interested in, because these parallel broadcasted multi-streams are not thematically interlinked.

In LIVE we will create these missing methods and tools for staging real-time parallel broadcasted video streams to create a multi-perspective show around one event.

4.1 Prior Research Work Concerning Design Principles and 'Vision Europe'

In our last IST-research project MECiTV we developed as video composers a first set of design principles to stage nonlinear iTV stories such as our docudrama 'Vision Europe'. Vison Europe is a nonlinear docudrama for iTV which leads the viewer through a staged story composition, a labyrinth of time and space. The viewer sits on a sofa in front of a TV screen, the remote control in her/his hand, and will follow individual people on the screen on their search for the Vision Europe. As soon as she/he is attracted by a protagonist, the

visitor can follow her/ him into her/his life, work, find more out about her/ him and take a travel through time with her/him. Just by pressing the O.K. button on the remote control in special moments in time the visitor can switch the video streams.

The visitor is immersed into the story and she/he will press intuitively, when she/ he feels like it. Sometimes she/he gets an explicit hint by the moderator on the screen, that it is possible to switch the video service, sometimes it is a second video screen or just a feeling. If the visitor feels satisfied with what she/he watches on the screen in the moment, she/he may not switch at all. There is no punishment for passive viewers.

The docudrama started on April 12th, 2003 in Budapest, when the Hungarians voted for joining the EU and the filming ended on May 1st, 2004, when the country finally entered the EU. Some video archive material used in the film covers the past, and the protagonists are talking about the future. It is a time travel through the past, present and future, stories of the past are linked to stories of the present and the future by reappearing protagonists and locations.
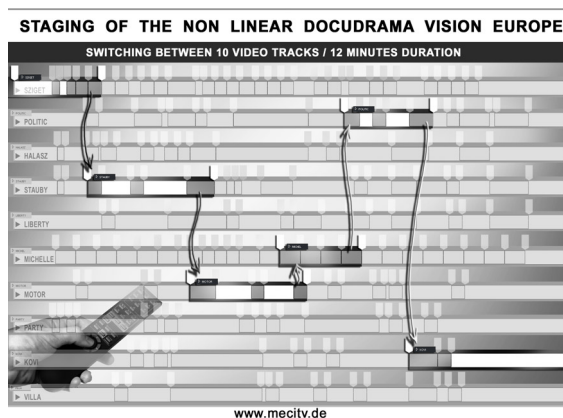

Figure 2: Staging of the Nonlinear Docudrama 'Vision Europe'

In Vision Europe ten different 'paths' of a story are told in parallel. Two tracks are called 'main streams' and form the structural backbone of the production (i.e. 'Sziget' and 'Michelle'). In Figure 2 the ten tracks are shown as horizontal bars. The time when the viewer is able to switch from one track to another is represented by a transition point. Thus the viewer is only able to switch during these intervals from the current track she or he watches to another by pressing the 'OK' button on the remote control. Pressing the OK button outside these intervals leads to one of the two main streams.

4.2 Content Based Approach

In MECiTV we proposed not to use any special graphical elements for a production or at least to use them only sparsely. Instead, we here state two design patterns which are content based approaches which stimulate audience interaction, namely: *Moderator Driven Navigation* and *Navigation by Visualisation*. These patterns were explored in the actual production 'Vision Europe', which was produced with the help of our iTV Composing Tool.

4.3 Moderator Driven Navigation

It is realised by a story path where the end user is directly addressed by a moderator. The moderator explicitly asks the user to use the OK button of the remote control to get for example
more information on a person. The moderator also explains the alternatives the user has. As an example, not using the switching functionality will lead to an automatic start of a certain story. This easy concept helps to slowly introduce the viewers to the capabilities of iTV and motivates interaction. On the other hand, the drawback of this method is that a moderator or at least a voice is needed to explain the interaction possibilities. Besides the increase of personal expenses, the tension of the global story or immersion of the viewer may be lost if many interventions of a moderator are made.
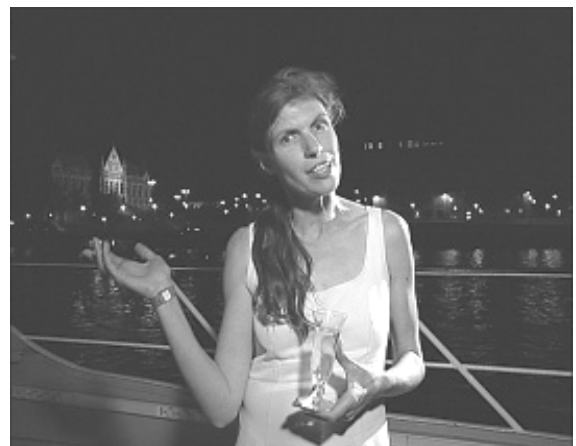

Figure 3: Guidance: The moderator tells the user explicitly to follow her by pressing the remote control

4.4 Navigation by Visualisation

The second pattern we explored is called 'navigation by visualisation'. Instead of the proactive interaction request by a moderator, the

viewer is seduced to switch by providing audio-visual temptations. The idea is to not break the story flow and the tension but instead give the viewer some subtle audio-visual clues on the actions in another story path. In our production this is achieved by attaching a small moving window showing the current scenery of another story path without sound (Figures 4-9).

In contrast to the first pattern, the viewer can already see what to expect if pressing the OK button and switching to the target path. The target path itself can again have reference to another target path and so forth. In Vision Europe we implemented three different variations of transitions: Seduction, Illustration and Temptation.

Figure 4 and 5: Seduction

For the transition *Seduction* the protagonist seduces the user by a body gesture to follow her.

Figure 6 and 7: Illustration

For the transition *Illustration* a second video illustrates the content of the main video and gives a deeper insight into the story.

Figure 8: Temptation

Figure 9: Temptation

For the transition *Temptation* a second video appears on the screen tempting the user to switch to another channel and change the experience. Among other this first set of design patterns will also be tested and used in our upcoming research project LIVE by the video conductor for staging of live media events.

5. Results of MECiTV

After its premiere for an invited audience at the Laboratory for Mixed Realities in 2004 the nonlinear docudrama 'Vision Europe' was presented at the IBC 2004 in Amsterdam and at the 2004 IST Conference in The Hague. In addition we created a special DVD version of the docudrama which was handed out to numerous individuals from the media production and TV industry around Europe.

People who watched and interacted with our 'Vision Europe' production for the most part liked the concept and understood it quickly. Only a very short introduction of two sentences was necessary to explain the interaction concept and the use of the OK button of the remote control. Reactions of viewers ranged from 'what's the novelty?' to 'very interesting'. Interestingly, out of the small percentage of people asking for the novelty in this concept many were expecting 'typical' iTV behaviour like graphical user interfaces and popup windows. Overall it is fair to say, that this prototype of an iTV production was a great success and people are ready to accept this kind of iTV production. It was even proposed by some to transfer this concept to other media such as broadband internet access (video on demand) and radio (interactive radio production).

As a result of the reactions several suggestions concerning the improvement of iTV production and application could be extracted. Envisioned formats that would go far beyond our prototype were the transition from only *pre-produced* material to the use of *live* material. Thus in MECiTV the central idea for the project LIVE was born, that the interactive coverage of live events – in particular elections and major sport events like the Olympic Games – would be the perfect playground for challenging new forms of iTV content.

6. Conclusions and Future Work

We have proposed a novel digital broadcasting framework, which allows for the live staging of (mass) media events and an improved consumer personalisation due to its evolutionary character. In addition new professions for future TV production workflows will emerge, namely the 'video composer' and the 'live video conductor'.

Starting from results achieved in a preceding iTV research project we will develop the missing methods and tools for the video conductor to stage live media event such as the 2008 Olympic Games in our upcoming IST Integrated Project (IP) 'LIVE'. Extensive research will be done to develop methods and tools to enable the professional user to link live video streams at special (spontaneous) points in time, so that the consumer can switch channels due to her/his interests.

A central question to be answered will be: How can the professional user stage content based and time-driven live video streams and get feedback from the consumers directly? 'Content based' means that the professional user can link videos when the content demands it. If for example an athlete wins unexpectedly, in a live moderation about the Olympics the video conductor can in this moment spontaneously link the winning video to a live interview with and (archived) background information about the athlete. In addition links might be established to other competitions in the same sport discipline out of the archive. The consumer is able to switch to one of these video streams which are thematically interlinked to the video stream she/he just watched. The linking of videos has to be time-driven, because the parallel streams are broadcasted and have to be synchronised due to time.

Finally we will do research on methods and tools to analyse and visualise the consumer feedback for the professional user in real-time, who needs answers to questions like: Does the consumers like the video show in progress or not (voting / explicit behaviour)

and how does she/he navigate through the offered streams (implicit behaviour)?

Our vision is, that the viewer of tomorrow will not follow channels anymore. Rather individual consumers will follow their favourite video conductor or even navigate by themselves through a mass media event such as the Olympic Games. At times the viewer will be in the mood to lean back on the sofa and be guided by the video conductor, then she/he leans forward and takes control over the journey through the video streams. Thus we will serve this basic human need of loosing control (being entertained, typical couch potato behaviour) as well as taking over the control (inform her/himself, typically PC behaviour).

References

[Cardinal] Cardinal Information Systems Ltd.: *Cardinal Studio,* Retrieved March 12, 2006, from http://www.cardinal.fi/studio/studio.html.

[Davis] Davis, M. (1995). Media Streams: An Iconic Visual Language for Video Representation*. In Baecker, R.M., Grudin, J., Buxton, W.A.S., & Greenberg, S. (Eds.), *Readings in Human-Computer Interaction: Toward the Year 2000* (pp. 854 – 866). San Francisco, Morgan Kaufman Publishers, Inc.

[Emuse] Emuse-Technologies. *Modelstream.* Retrieved March 12, 2006, from http://www.emuse-tech.com, Dublin.

[Ensequence] Ensequence: *on-Q Create Suite.* Retrieved March 12, 2006, from http://www.ensequence.com/products/on-q_create/, Portland, Oregon.

[Freeman] Freeman, J, & Lessiter, J. (2003). Using attitude based segmentation to better understand viewers' usability issues with Digital and Interactive TV. In Masthoff, J., Griffiths, R. & Pemberton, L. (2003). (Eds.) *Proceedings of the EuroITV03,* Brighton, UK.

[Lee] Lee, Hee-K., Lee, Han-K., Nam, J., Bae, B., Kim, M., Kang, K. & Kim, J. (2002). Personalized Contents Guide and Browsing based on User Preference. In *Proceedings of the AH2002 Workshop on Personalization in Future TV,* Malaga, Spain.

[Melinger] Melinger, D.J. (2004). *Massive Media,* Thesis, Interactive Telecommunications Program, Tisch School of the Arts, New York. University. Retrieved March 12, 2006, from http://massive-media.org/docs/m-m.pdf.

[Molander] Molander, P., & Wolf, S. (2002). *Case Study: Automation of the BBC's interactive TV playout system.* IBC Conference Papers. Retrieved March 12, 2006, from http://www.broadcastpapers.com/data/ibc2003ITV BBCInteractTV.pdf.

[Oncotype] Oncotype Production (2003). *Switchin.* Retrieved March 12, 2006, from http://www.switching.dk/en.

[Rhambia] Rambhia, A., Wen, G., & Cheng, S. (2004). Content Morphing: A Novel System for Broadcast Delivery of Personalizable Content. In Ardisson, L., Kobs, A., & Maybury, M. (Eds.), *Personalized Digital Television* (pp. 235 – 255), Human-Computer Interaction Series 6, Kluwer Academic Publishers.