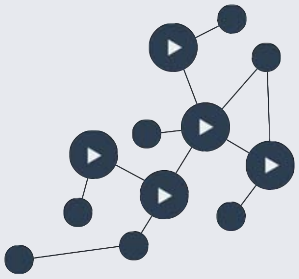
2nd International Workshop on Interactive Content Consumption at TVX 2014

25th June 2014, Newcastle upon Tyne, UK

Proceedings



tvx 2014

Introduction

At this year's ACM TVX conference, the second edition of the International Workshop on Interactive Content Consumption (WSICC'14) took place in Newcastle upon Tyne, UK.
It drew considerable interest, with about 30 participants over the course of the full day event.

The workshop's aim was to shed light onto the latest developments in the research landscape behind the workshop's scope, and to find ideas where interdisciplinary institutions could aspire collaboration. It addressed a balanced community from industry, technical and social sciences research.

WSICC's keynote was held by Vconect's scientific coordinator Prof Marian F. Ursu (University of York) on the subject of 'Intelligent Video Interaction'. The talk highlighted several research projects on the topic. He outlined how advancements regarding different aspects feed into each other. In his keynote, Marian provided a well-reflected view on the interplay between handcrafted video narratives and semi-automated technology that supports broadcasters in deploying large scale interactive video projects. This both insightful and entertaining talk triggered lots of interest, follow-up-questions and discussions from and amongst the audience.

The full day workshop format further consisted of research paper based talks, a poster and demo session and two rounds of fishbowl discussions. The demo session was kicked of by the authors explaining the basic idea of their submission in one minute on one slide (madness session). A fishbowl discussion is an interactive discussion format that already proved to be very successful in the first edition of WSICC (2013 in Como, Italy).
As expected, the workshop participants embraced the fishbowl discussion format very quickly and engaged in intense and insightful discussions. While the debates on current forms of interactive content content consumption and future challenges went on, participants were invited to take notes and add them to a concept map. This provided a useful documentation of the discussions and allowed to evaluate the findings, questions and ideas after the workshop, as well as to merge them with the results of the first WSICC edition.
WSICC was organized by Britta Meixner and Katrin Tonndorf (University of Passau, representing the mirKUL project), Joscha Jaeger (Merz Akademie) and by Rene Kaiser of JOANNEUM RESEARCH, representing the Vconect project. Overall, the workshop can be considered very successful.
Preparations for another edition are already ongoing.

Enjoy reading WSICC's proceedings,
Britta, Joscha, Katrin and Rene

Further information:

Workshop:
http://wsicc.net/2014/

Conference:
http://tvx2014.com/

Print Proceedings:
http://wsicc.net/2014/Proceedings-WSICC-2014.pdf (open access)

Online Proceedings:
http://wsicc.net/2014/#proceedings

Visual impressions:
https://www.flickr.com/photos/49520289@N04/sets/72157645061214918/

Call for Papers

2nd International Workshop on Interactive Content Consumption at TVX 2014

Britta Meixner, Katrin Tonndorf
Passau University, Germany
meixner@fim.uni-passau.de
katrin.tonndorf@uni-passau.de

Rene Kaiser
JOANNEUM RESEARCH
Graz, Austria
rene.kaiser@joanneum.at

Joscha Jaeger
Merz Akademie Stuttgart
Stuttgart, Germany
joscha.jaeger@merz-akademie.de

ABSTRACT

The second edition of the WSICC workshop focuses on novel forms of interactive content consumption. It will explore the shifting balance between lean-back passive TV and Web-based media consumption, and lean-forward interactivity. This shift is especially relevant considering new forms of collaborative content creation, controlling the media with a companion screen, and more advanced forms of audiovisual content interaction. Therefore, new media formats and consumption paradigms have emerged that allow for new types of interactivity. Beyond entertainment, interactive audiovisual content has strong potential for learning and support scenarios.

Author Keywords
content consumption, interaction, multimedia content

ACM Classification Keywords
H.5 Information Interfaces and Presentation: Multimedia Information Systems

WORKSHOP AIM

The workshop's objective is to provide a highly interactive discussion forum that allows capturing a comprehensive view on the research area it addresses. During the workshop, an overview on new content interaction concepts, research activities and future challenges in this area will be concluded and documented. An interdisciplinary view on the topic shall be compiled by contributions from technical research, conceptual work, user-centric studies, industry developments, as well as experimental showcases. Part of the discussions are fueled by technical demonstrations of interactive content consumption forms. The workshop aims to examine and evaluate new forms of content interaction by discussing the field along three axes:

- Recent technological advances that provide new forms of audiovisual content interaction;
- Technologies that supports users in finding the balance between passive consumption and lean-forward interaction;
- User studies that evaluate new types of audiovisual content interaction.

The following will sketch areas and aspects that are considered within the scope of the workshop. We seek technological research on more active interaction with audiovisual content, e.g. collaborative hypermedia generation, social media integration, game-like interfaces, or intelligent storytelling and narrative engines. The workshop deals with both recorded and live media access. Mobile and domestic consumption may be investigated. Topics are not limited to entertainment and learning, but should remain within the overall scope of the workshop. Below, some of the questions that the workshop aims to answer are listed:

- How can forms of (inter-)active media access be designed to be interwoven with passive consumption modes?
- How does the balance between active and passive consumption affect the Quality of Experience?
- How can interactivity enhance the experience of people watching together, even when they are in disjoint locations?
- How can active and passive content consumption foster learning?
- How can content personalization be enhanced through interactivity, and at which abstraction level do users want to interact?
- What do recent studies on interaction with content in the realm of social media sharing reveal?
- How can we understand how users want to use physical devices to interact?
- Which requirements for content production do new forms of interactive media imply?
- Do trends in content consumption behavior influence technical research by revealing new challenges?
- Which technical advances are needed to allow the industry to offer more interactive media services?

The workshop accepted 3 types of submissions: full research papers for presentations, short papers for poster oral presentations, and technical demos. The program is available on the workshop's website, http://wsicc.net.

WORKSHOP FORMAT

We have developed a workshop format to stimulate networking and knowledge transfer among the participants. The full day workshop will be an active forum to discuss

TVX'14, June 25 – June 27, 2014, Newcastle, UK.
DOI.10.6084/m9.figshare.1031601

research challenges, methodologies and results in a field that is gaining attention based on quickly changing content consumption needs and habits. More than half the time will be reserved for discussion. The chairs will establish an informal atmosphere, inspired by Barcamps. In an active moderating role, they will make sure the workshop's questions will be answered and documented, yet will allow some flexibility where appropriate to meet the interest of the audience. Results will be collected on flip charts along multiple questions which emerge during the day, e.g. what are the latest innovations in that field? Which research activities exist to tackle unsolved challenges? How could we combine different interaction technologies to the benefit of the user?

Throughout the day, the audience will be encouraged to contribute, and especially to comment existing inputs. The outcome of the highly interactive part of the workshop will be summarized on a poster for the presentation at the main conference. WSICC will consist of the following sessions:

- Welcome and presentation of workshop aims.
- Interactive participant introduction in Barcamp style (name, affiliation, role, 3 keywords/hashtags).
- Invited keynote by Prof Marian Ursu.
- Pitches (a 2min) to kickstart the poster/demo session, where, in parallel, posters and demos are exhibited and discussed. This shall establish an understanding of each other's work, approach, and focus.
- Three research paper based talks. Questions will be allowed during the talks.
- A fishbowl discussion format, around aspects raised during WSICC. There is a limited number of active seats. If you want to say something, you have to take an empty seat or wait for one. This format of a dynamically changing working panel proved to work well for discussions among experts on concrete questions.
- Concluding session. The group will revisit what has been collected throughout the day. Conclusions will be summarized.

ORGANIZING COMMITTEE

The organizing committee consists of 4 members who cover different areas of technical research, HCI research and communication research.

Britta Meixner is a researcher at Passau University. She received a diploma in computer science and a state examination for lectureship at secondary schools from the University of Passau, Germany, in 2008. Currently, she is working towards a PhD degree in computer science at the Faculty of Computer Science and Mathematics of the Passau University. There, she is conducting research and development in the area of hypervideo. On the one hand, she focuses on creating easy to use authoring tools and players and on the other hand download and cache management are evaluated to provide a better viewer experience. Further she is interested in hypervideo on mobile devices, collaborative hypervideo creation, and decision rules in hypervideo. Britta is a member of the BMBF research project “mirKUL” that investigates application scenarios of interactive nonlinear video. Britta is a reviewer for the Multimedia Tools and Applications Journal (Springer) and was a member of the program committee of the 1st International Workshop on Interactive Content Consumption at EuroITV 2013.

Katrin Tonndorf is a researcher at Passau University. She received a magister degree in media studies from the Technical University Braunschweig and the Braunschweig University of Arts in 2010. Currently, she is working towards a PhD degree in communication studies at the Faculty of Arts and Humanities at the Passau University. She is conducting research in the area of online and social media communication practices. Furthermore she is interested in the use of interactive audiovisual content for learning und support purposes. Katrin is also a member of the BMBF research project “mirKUL”.

Rene Kaiser is a key researcher for JOANNEUM RESEARCH and has been involved in a number of European projects dealing with automation of content production such as NM2, APOSDLE, TA2 and Vconect. His research focus is on *Virtual Director* software, on automating shot selection through cinematographic behavior models. Further he is interested in automating non-linear video production, enabling the user to interactively influence the narrative path while watching. Rene was responsible for the organization of the Interactive and Immersive Entertainment and Communication Special Session at MMM'12. He is part of a group hosting the annual PhD cooperation workshop at the i-KNOW and i-SEMANTICS conferences, active member of IEEE STCSN, and co-organizer of the Barcamp Graz, a yearly 3-day *unconference* which is an interactive and open discussion format. At EuroITV 2013, Rene was co-organizing the first edition of WSICC.

Joscha Jaeger is a research assistant at Merz Akademie Stuttgart and founder of filmicweb - Hypervideo Interface Design. His research covers Web-based hypervideo technology, time-based interaction, and semantic video search interfaces. Joscha has a strong focus on film as information architecture, collaborative editing systems for non-linear film, and user-driven annotation systems. He is interested in finding new ways of distributed interaction with open video technologies and interfaces on the web.

ACKNOWLEDGMENTS

The workshop is supported by the European Commission under the contract FP7-287760, “Vconect - Video Communications for Networked Communities” and by the German Federal Ministry of Education and Research (BMBF) under contract 03V0633.

Posters / Demos

Using Overlays and Second Screen to Stimulate Social Interaction Without Compromising Passive Consumption

Rinze Leenheer
CUO, iMinds|KU Leuven
Parkstraat 45, 3000 Leuven Belgium
rinze.leenheer@soc.kuleuven.be

David Geerts
CUO, iMinds|KU Leuven
Parkstraat 45, 3000 Leuven Belgium
david.geerts@soc.kuleuven.be

ABSTRACT

Creating a good interactive TV experience that does not compromise the 'regular' TV experience is no simple task. We propose a combination of second and first screen to offer a social-interaction stimulating, interactive experience that is completely optional. First observations and interviews have been completed. The next steps are a series of prototypes and a field pilot test in the Netherlands.

Author Keywords

Interactive TV; Second Screen; User Experience; Social Interaction

ACM Classification Keywords

H5.1 [Information Interfaces and presentation]: Multimedia information systems – audio, video; H5.m [Information Interfaces and presentation (e.g., HCI)]: Miscellaneous

INTRODUCTION

Interactive forms of TV have been explored for a long time. In the beginning the main focus was on interaction on and through the main television screen [5], this has shifted to interaction on and through second screen devices [1].

With the rise in popularity of smartphones and tablets the amount of commercial second screen applications has also increased rapidly. However according to app builders and broadcasters we interviewed, the percentage of TV viewers that actually use these apps is still quite low [4]. Therefore it is important to them that these second screen applications do not compromise the 'regular' TV experience.

Interweaving the active and the passive

TV is traditionally a lean back medium and many people use it this way to relax. Care should be given when adding interactivity to not estrange these passive viewers. The way forward lies in a method to deliver a rich interactive experience to the part of the audience who wants it, that does not compromise the passive viewing experience.

Social Interaction During TV Viewing

Social interaction can enhance the TV experience when applied to the right TV genres [3][6]. There are two forms of social interaction for TV collocated and remote. There has been quite a substantial amount of work done for remote social interaction [2][3]. But less that has focused on the collocated interactive experience. Obviously viewers in the same room can already communicate with each other but it would be interesting to see how second screen devices can be used to stimulate social interaction in the living room.

2nd International Workshop on Interactive Content Consumption at TVX'14, June 25, 2014, Newcastle, UK.

LESSONS LEARNED FROM OBSERVATION AND INTERVIEWS

We previously investigated a commercial second screen companion application that offered extra information in accompaniment to a television program. Based on observations and interviews with viewers and producers, and actual usage data of a companion app from Google Analytics, we discovered several insights and recommendations on how to design companion apps related to ease of use, timing, social interaction, attention and added value [4].

One of the things we learned is that viewers enjoy the added layer of interactivity to their 'passive TV show', however the 'interactive' viewer is usually still just a very small part of the total audience. Broadcasters and Program makers are therefore hesitant to implement interactivity that could 'bother' these 'regular' viewers. As one broadcaster put it:

"The problem is that many program makers say the group of second screen users is only a small percentage of our viewers and won't change their show just for them."

One way to offer an interactive experience, that viewers who do not want to participate will not notice, is through the use of a second screen. A second screen application can offer extras without interfering with the first screen. Because second screen devices are usually personal devices, like tablets and smartphones, even when there are several people watching a TV program in the same room, one or more of them can participate in the interaction without disturbing the others who do not wish to participate. The market penetration for second screen devices is also high. Therefore availability should not be a concern.

Another concern from TV producers is that a second screen companion app will distract its users from the television show itself.

"A challenge is to convince program makers that the second screen won't distract from the first screen experience." – Second screen app developer

From the observation and interviews we found that viewers that use these second screen apps use non program related second screen applications like email, Facebook or Twitter when there is no dedicated second screen application available for the show itself.

"If there's nothing happening for a while I tend to switch to something else. You might miss some updates then, because you don't switch back to the app in time."

Dedicated second screen applications could therefore actually increase the attention of these users for the TV program. As a second screen app developer puts it:

"If you offer a good second screen app with a TV show and you can engage viewers through this second screen, you will end up with a more attentive TV viewer."

In conclusion using a second screen to deliver an interactive experience for a TV show has the advantages of having a readily available infrastructure and the possibility to not disturb people who just want to passively enjoy the show. There is however also a big disadvantage of relying solely on the second screen for interactivity.

MOVING ON WITH FIRST AND SECOND SCREEN

The downside of the use of a second screen for interactivity is the loss of a group feel when watching a show together in the same room. Everybody uses an individual device. It is possible for people to share a device but this brings with it its own problems with polls or quizzes because only one answer can be given. The TV screen is the central device in this scenario. The difficulty is using the TV to facilitate a group experience without compromising the experience of those who do not want to participate in the interactive portion of the TV show. A way to solve this is by using overlays on the TV. Overlays are a way to show information to a TV viewer that is not imbedded in the broadcast signal. Overlays are used to show TV viewers the EPG or TV menus for instance. With the use of overlays in combination with second screen devices, it is possible to offer a group experience that doesn't interfere with the show itself while still giving everyone control through their own device.

To find out how a setup like this impacts group dynamics, we are planning a pilot with an interactive quiz application as an addition to the Dutch TV show 'De Rijdende Rechter. This pilot will be executed as part of the TV-Ring European research project. "De Rijdende Rechter is a Dutch TV show that deals with a judge ruling in disputes between civilians. There is a quiz element to this show in which viewers can predict how the judge will rule on certain statements. At the moment anyone with a second screen device can participate but there is no way to see scores or a status from other participants. In the pilot an overview of the participants and their scores will be presented on the TV through an HbbTV overlay. We predict that this will lead to more social interaction between the people who watch and participate in the quiz together and thereby to more enjoyment. In the first prototyping phase we will create different versions of the quiz. This will be a round of paper prototypes. The goal of this phase is to try many different methods for different phases of the game, like the setup, the questions, and the scores, to investigate which combination stimulates social interaction the most. The methods will differ in the amount of information (names, scores, rank, answer status etc.) that is presented and on which screen this information is displayed (TV, second screen). After the initial phase the most promising methods will be turned into interactive prototypes. The end goal is to have a live version that will be fully functional which will be developed in collaboration with our TV-Ring partners from the NPO (Nederlandse publieke omroep, www.npo.nl) and Peoples Playground (www.peoplesplayground.nl).

DISCUSSION

Interactive TV applications are very much enjoyed by a certain group of users. However this group is still relatively small compared to the whole of the TV audience. Because of this Broadcasters and TV makers are hesitant to invest much in interactive TV development and they certainly do not want to adapt their regular programs too much and risk alienating this majority of passive viewers. This leads to a chicken and egg conundrum: because TV makers are hesitant to offer interactive content, the state of interactive TV is not developing very rapidly which in turn leads to the group of interactive viewers staying small which leads to little investment in interactive TV etc.

We already found second screen applications to be a good way to offer an interactive TV experience that does not interfere with regular viewing. However the second screen alone does not maximize the increased social interaction interactive TV can offer. We propose a setup with second screen and overlays on the first as a solution. This setup is able to deliver a rich interactive experience that stimulates interaction to the viewers who seek it out while at the same time being almost invisible to the viewers who want to 'simply watch their show'. We believe it can play a big role in solving the interactive TV chicken and egg conundrum.

ACKNOWLEDGEMENTS

The research leading to these results was carried out in the TV-Ring project (EC grant agreement ICT PSP-325209).

REFERENCES

1. Cesar, P., Bulterman, D.C.A., and Jansen, A.J. Usages of the secondary screen in an interactive television environment: Control, enrich, share, and transfer television content. *Changing television environments*. Springer Berlin Heidelberg, 2008. 168-177.

2. Chorianopoulos, K., & Lekakos, G. Introduction to social tv: Enhancing the shared experience with interactive tv. *Intl. Journal of Human–Computer Interaction*, *24*(2), (2008), 113-120.

3. Geerts, D., & De Grooff, D. Supporting the social uses of television: sociability heuristics for social TV. In *Proceedings of the SIGCHI conference on human factors in computing systems,* ACM Press (2009), 595-604).

4. Geerts, D., Leenheer, R.A., De Grooff, D., Negenman, J., Heijstraten, S., In Front of And Behind The Second Screen: Viewer and Producer Perspectives on a Companion App. *Accepted TVX 2014*, (2014)

5. Jensen, J.F., Interactive Television: New Genres, New Format, New Content. In *Proceedings of the Second Australasian Conference on Interactive Entertainment*. Sydney, Australia, Australia: Creativity & Cognition Studios Press (2005), 89–96.

6.Sperring, S., & Strandvall, T. Viewers' experiences of a TV quiz show with integrated interactivity. *Intl. Journal of Human–Computer Interaction*, *24*(2), (2008), 214-235.

Location Based Video Flipping: Interactive Prototype navigated by HbbTV remote control

Thomas Fritzsche, Stefanie Müller, Arne Berger, Maximilian Eibl
Technische Universität Chemnitz
Strasse der Nationen 62, 09111 Chemnitz, Germany
firstname.lastname@informatik.tu-chemnitz

ABSTRACT

We present a geospatial navigation concept for browsing videos according to their tagged geographic location. The proposed application is derived from two modi operandi for selecting video content: While continuous, yet one-dimensional flipping through listed video clips can be controlled with a regular HbbTV remote control, the discrete selection of video clips that are positioned on maps according to their tagged location is usually done with mouse and cursor. The proposed concept combines the ease of use of a remote control in a lean-back setting and the precision of selecting videos on maps.

Author Keywords

Devices & Interaction Techniques, HbbTV, Location Flipping, Location Based Services, YouTube API, Remote Control, Lean Back Setting

ACM Classification Keywords

H.5.m. Information interfaces and presentation

INTRODUCTION

Combining the ease of use to navigate small amounts of video clips with simple navigation devices with the benefit of accessing huge numbers of clips is a challenging task [1]. We introduce the term "location flipping", which stands for toggling through videos on geographical maps while benefitting from the ease of use of a remote control sufficient in lean back settings. Universally known flipping corresponds to hopping through lists of television channels. In more interactive settings as in Web TV users are able to choose from videos that are positioned according to the videos geospatial location on a geographic map. However, the ease of use of a remote control is not provided. Thus, location flipping aims to combine thonse both paradigms for navigating geospatial content with a regular HbbTV remote control in a lean back setting. Since standard HbbTV remote controls are equipped with a four way navigation and four colored buttons for additonal functions, they will be used for the proposed application.

We present the prototype of a novel interface for lean back remote control navigation and selection of videos on maps and the corresponding implementation for retrieving video clips depending on their geographic content on a map.

2nd International Workshop on Interactive Content Consumption at TVX'14, June 25, 2014, Newcastle, UK.

RELATED WORK

Flipping through a finite variety of channels on a TV set is a well-known task during habitual video browsing. Here, a remote control offers few buttons and is favourable for that kind of lean back environment. In contrast, lean forward settings allow the navigation of an infinite amount of potentially more suitable video clips. However, a more advanced navigation and consequent user involvement will be necessary [2]. The already existing approaches for selecting video content that are presented on maps with the use of a remote control lack among other things the feature of video previews.

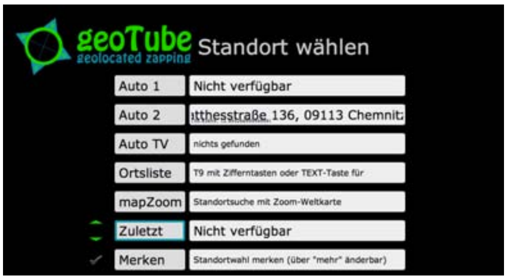
USER CENTERED DESIGN PROCESS

To determine user expectations of location based zapping and generate insight for developing this novel lean back interface, first a questionnaire was conducted. While gaining insight into user behaviour, adequate participants for the following workshop were recruited. Subsequently an analog prototype was developed and evaluated. Following the workshop the here presented interactive prototype was implemented and evaluated.

USING A HBBTV REMOTE CONTROL

To navigate interactive items an a TV screen, most applications use the remote controls four way navigation consisting of four cursor keys grouped around a center button as well as four colored keys. Also the workshop proposed navigating the four cardinal directions with the four cursor keys on the HbbTV remote control.

INTERACTIVE PROTOTYPE



Standort wählen

Auto 1	Nicht verfügbar
Auto 2	ıtthesstraße 136, 09113 Chemnit:
Auto TV	nichts gefunden
Ortsliste	T9 mit Zifferntasten oder TEXT-Taste für
mapZoom	Standortsuche mit Zoom-Weltkarte
Zuletzt	Nicht verfügbar
Merken	Standortwahl merken (über "mehr" änderbar)

Figure 1. Digital Prototype – Selecting Locations

The application is started by the menu of the receiver. The text input is realized by the receivers remote control numeric keys. To ensure the readability of the font, overlong text is scrolled like a ticker.

The application offers a variety of modes to select a location. After an initial location is selected, surrounding clips can be navigated by the remote controls four direction pattern (figure 1).

- *Auto 1 and 2*: With the help of the receivers IP adress and two free usable services, the geographic position of the viewer is detected by a database. Any queries will only being send in case of a changed IP adress.
- During a continuous TV program, the function *Auto TV* will search for specified words in the description of the current watched show.
- The *List of locations* enables the search of available locations through text input via an on screen virtual keyboard.
- The active selection of locations and zooming in a map by numeric keys is available by choosing the function *mapZoom*.
- The function *last* makes the latest used geoposition available.
- Users are able to *save* their current geoposition within the application by selecting the function.

After choosing a location, the main view of the application is started (figure 2). The video clip, that is nearest to the selected geographical position, is provided as the biggest thumbnail in the middle of the screen. This video can be scaled to full screen mode and a provided timeline enables the user to fast-forward and rewind. This centered video clip is surrounded by four more video previews in the north, south, east and west attached to four triangles representing the cardinal directions. Those triangles are marked with the words north, east, south and west to increase their meaning and can be accessed by the remote controls arrow keys. In addition, two more but smaller thumbnails are positioned above and below of each cardinal one, which display a preview of the subsequent video clips. The shown previews in north, south, east and west represent the nearest video clips to the selected clip and are overlayed by their titles. More functions can be accessed with the four colored menu items present in the lower third of the screen, that are refered to by the four color buttons on the HbbTV remote control. By pushing the red button, a navigable map of the currently picked area is shown. The green key offers several filtering options for the retrieved video clips and the yellow one a list of all search results. The blue button is reserved for yet to be determined features like favorites, location selection or other informations.

FUTURE WORK

The goal of the proposed interactive prototype was to determine if users find the idea of location zapping useful and whether our proposed navigation concept for flipping location based video clips is suitable for that application. Indeed, the concept is efficient and feasible, however a variety of future functionalities has to be considered due to the constraints of a HBB TV remote controls. Although, parameters like the search distance threshold, duration, creation date as well as search and input of a position need to be included.

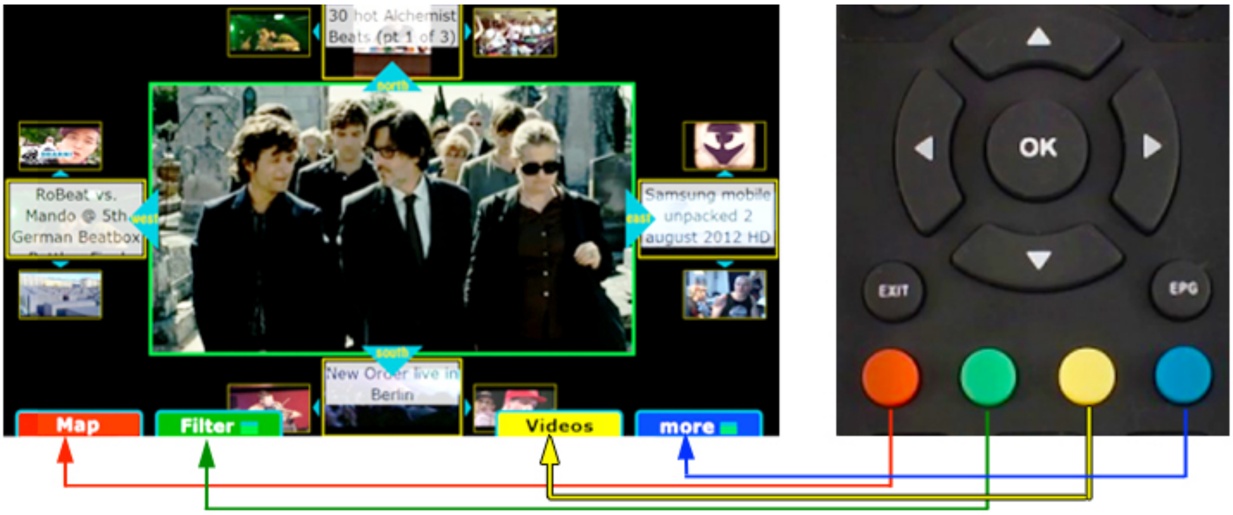

30 hot Alchemist Beats (pt 1 of 3)
north
RoBeat vs. Mando @ 5th German Beatbox
west
Samsung mobile unpacked 2 august 2012 HD
east
south
New Order live in Berlin
Map
Filter
Videos
more
OK
EXIT
EPG

Figure 2. Digital Prototype - Main TV-view

REFERENCES

1. Knauf, R. et al. 2010. Constraints and simplification for a better mobile video annotation and content customization process. *Workshop Proceedings of the EuroITV*.
2. Berger, A. et al. 2011. Moody Mobile TV: Adding Emotion To Personalized Playlists. Proceedings of the Mobile HCI Conference.

HbbTV based Augmented Information Television with Segment-linked Related Content on TV and 2nd Screen

Robert Strzebkowski **Roman Bartoli** **Sven Spielvogel** **Danilo Schmidt**

Beuth University of Applied Sciences for Engineering, Luxemburger Str. 10, 13353 Berlin

robertst@bht-berlin.de roman.bartoli@gmail.com spielvogel@bht-berlin.de an.danilo.schmidt@gmail.com

ABSTRACT

In this paper, new options are presented for seamless connection and synchronization between linear – passive / lean back - Television broadcasts and interactive – lean forward - online multimedia content as well as 2nd Screen applications. Based on two project examples it will be shown and solutions discussed concerning scene- & segment-based synchronization between the content of information/news/documentary or children TV broadcasts and related additionally 'stretched' online content as well as interactive applications – both on TV and/or on 2nd Screen. The projects and examples are based on the emerging interactive TV standard HbbTV – Hybrid Broadcast Broadband TV. A mixed passive / interactive consumption approach is an important issue here.

The first example is a very well-suited solution for the TV program categories 'news', 'information' or 'documentary' and thus for *informal learning scenarios* involving mixed TV/Internet solutions. This is a joint project with Germany's public broadcaster ZDF for the daily 30-minute news magazine 'heute-journal'. The aim was here to provide viewers with access to synchronized complementary information, simplifying broadcasting techniques and facilitating an intuitive way and non-disruptive form of interactivity and presentation. Other related content and Apps are to be used as flexibly as possible, either on the TV screen, a connected second screen device, or in a parallel manner on both devices. The technical basis for this is HbbTV.

The second example is an actual technical concept study implemented for interwoven TV and 2nd Screen content and application. Firstly, the TV-based HbbTV application triggers the 2nd Screen app to change the presentation, content and functionality of the app. We call this effect 'Application Triggering'. Secondly, there are different content and application types focusing either on television or mobile devices that mutually complement each other – we call this approach 'Splitting Content & Splitting Application'.

This second example will demonstrate the extensive interactivity vis-a-vis creative visual tasks. The 'viewser' should A) directly address a specific topic and B) promote their creative work for the viewer community. The application also allows simultaneous / collaborative work on a mutual creative 'project' and the joint presentation of the results to the public.

The aim of such new technical, interactive, cross-media/-channel and cross-device solutions is to promote *'Augmented Informal Learning with Television-based Media'*.

Author Keywords

Interactive Television, Smart-TV, Hybrid TV, Cross-Device, Second Screen, HbbTV, Connected-TV, Content-Enrichment, Synchronized Content, Informal Learning

ACM Classification Keywords

H.1.2 User/Machine Systems ,H.5 Information interfaces and presentation, H2. Database Management

INTRODUCTION

We are presently observing certain important developments in media devices infrastructure in terms of the form of media distribution, TV content presentation form, TV formats, and media usage patterns:

- Exponential increase in the usage of WebTV, Internet-Video and IPTV – both live and on demand. [1]
- Growing number of Internet-connected Smart-TV/Hybrid-TV (HbbTV) devices.[4]
- Growing number of complementary multichannel TV-based and TV-related content[1],[5]
- Wide range of daily information/ documentary/popular science programs being broadcast.[5]
- Exponential increase in number of mobile devices in particular tablets and 'smartlets'.
- Increased usage of mobile devices used to view television broadcasts – the emerging 2nd Screen phenomena.[1]
- Growing social exchange while, or associated with viewing a certain TV event or broadcast.[1]

2nd International Workshop on Interactive Content Consumption at TVX'14, June 25, 2014, Newcastle, UK.

Such developments not only prompt new consumption patterns and needs by the viewers of Television and Television-based media, they also give TV media producers and broadcasters/providers new opportunities in terms of new program formats and possibilities.

The Television sector already has access to novel new forms of interactivity and complementary content solutions. Content can be now appropriately distributed among the various media presentation formats, media channels and media devices, which ensures the most effective media impact and creating stronger bonds between the viewer/user 'viewser' and a specific certain media program/channel/.

Within the scope of 'Cross-Media' and 'Connected TV' approach there are also new opportunities to enrich content and to more closely engage viewers with the provided information. Particularly the interplay between large TV Screens and mobile second screens on Tablets or Smartphones is creating an 'Augmented Television' effect. It means the consumption possibility of complementary, additionally content to the running/playing broadcast – whether live or on demand.

How WebTV and IPTV impact traditional TV consumption

There is no doubt for the still exponential growing Internet based Video and Television content and its usage almost of each Internet user worldwide.[5]

The increasing use of 'Online-Video' also impacts cross-media usage patterns and behavior[7]. Online users are used to instantly and directly accessing their desired video/TV. Linear TV consumption is decreasing significantly among younger viewers in favor of active, time-shifted and on-demand TV content.[5] The interactive and non-linear approach to consuming media or any online content is often based on the ability to setting 'bookmarks', 'marking' or to recording certain IPTV content for subsequent viewing.

Hybrid-TV is the foundation for successful Interactive TV

There number of commercially available Smart and Hybrid TV devices is growing rapidly, increasingly with direct internet connectivity features.[4] However, there is a subtle fine difference between the conventional Smart TV usage based on installed TV Apps on the TV device, and the genuine 'hybrid functionality' of Smart TV as provided by the interactive Television standard HbbTV (ETSI TS 102 796).

Traditionally Smart TV Apps will be used asynchronous to the current running or on-demand TV broadcast/content. Typical applications here include video-on-demand portals, 7 days catch up Television, news, weather or multimedia magazines from diverse publishers.

The 'real hybrid' TV approach means a synchronous connection between the on-screen, on-demand or live TV image and an appropriate TV App or related content. Today we can already cross-link an exact scene or segment on the TV screen with additional online content or/and Apps. From the perspective our research group, this time and content based synchronization between the TV broadcast and related additional content or applications is a very important prerequisite in terms of ensuring the effective functionality of interaction between the passive and active TV related content.

Parallel viewing of TV & Internet with 2nd Screen

There are indications that the parallel use of Internet content and services during the consumption of TV has been increasing over the past few years.[7] Parallel use of Tablets and Smartphones as so called 2nd Screen devices during the TV consumption and in particular Tablets (and 'phablets' or 'smartlets' as small Tablets) has increased over the last two years and is increasingly playing a key role due to the enhanced usability and multimedia potentials these devices provide vis-a-vis conventional smartphones.[1]

Inasmuch as the 2nd Screen trend remains a relatively new phenomenon there are many different definitions and categorizations of the applications. The '2nd Screen Society' distinguishes between three major categories of 2nd Screen applications involving slightly different impacts on media experiences: A) Companion Apps (that complete the viewing experiences – e.g. Tweets to the TV event), B) Converged or Enhanced Viewing Apps (for the 'Momentum' Experience – e.g. current statistics during a live sport event) and C) Viewing Apps (for the Viewing Experience – e.g. Video Library of a broadcaster).[1] Increasingly, these three App categories tend to merge. Our projects are positioned between the Companion and the Converged categories.

The question then arises as to age group differences: which is the 1st or the 2nd Screen of choice? Statistics generally show that young people (14 – 29 years old) are still watching TV content, but usually not on conventional TV screens; this group prefers Laptop/PC, Smartphone or Tablet as their 1st Screen devices.[5]

The next important aspect regarding 2nd Screen usage is the issue of WHAT do 'viewsers' consume on their mobile devices while watching TV? An important survey conducted in Germany in 2013 shows that the most popular activities (multiple answers were possible) were checking and writing e-mails (65% by Smartphone/74% by Tablet) surfing on the web (48%/58%), visiting social networks (43%/46%) and more importantly for our project - (26%/42%!) are searching for additional content related to the currently viewed TV broadcast, 41% of Tablet users are searching for information about products offered in TV Ads, and 26% are using TV-related social networks.[10]

With more than 41% of the 2nd Screen activity related to searching information either about on-screen TV broadcast content or Ad content, it can be claimed that a significant number 'viewsers' are interested in obtaining synchronous information related to their current on-screen TV content

Information, Popular Science and Infotainment on TV

Statistics regarding the consumption of and offered TV in both Germany and in several other countries reveal a broad range of information-related broadcasts. In terms of genre, information-related broadcasts hold first place in Germany itself (the German language also extends to Austria and parts of Switzerland). In terms of consumption, this genre holds 2nd place, only 3% behind the popular fiction genre. The information genre also appears to interest young people, as it holds 2nd place after fiction and ahead of entertainment, translating to nearly 25% of the entire TV genre range.[3]

'News' is the most demanded information format and TV remains one of the main information sources among different media options. Approximately 80% of younger people, i.e. 'Digital Natives'[1] (average age approx. 25 years) visit at least one news option every day, with 90% of the 'Digital Immigrants' and more than 70% of 'Millennium Teenagers' also interested in daily news services. In Germany, the primetime news show 'Tagesschau' presented on the first public broadcast channel ARD attracts nine million viewers each day, and approximately 4 million viewers for the 30-minute news magazine 'heute-journal' presented by the second public broadcast channel ZDF. Television is still the chief information source and medium for news for the 'younger' audience (up to age 44) with approximately 31-35% using TV and 40% using Internet; for the 'over-44' audience the ratio of TV/Internet is 50% / 20%. [3][5]

INFORMAL LEARNING VIA INTERACTIVE HYBRID-TV

Television possesses enormous potential in terms of visualization and story-telling (drama), indeed an approach that can present information in a very dedicated manner. In recent years, a noticeable change in the visual form and presentation quality of information-related TV productions has been observed. In addition to new optical techniques such as tracking and flying camera shots at difficult locations such caves or water falls, there are mixed animated 3D imaging and FX techniques video shots able to visualize hidden structures in underground constructions or bridges for instance, particularly in the documentary sector.

For so called 'fictional documentaries' – Docufiction – theatrical and dramaturgical elements are increasingly used and mixed with 'real' documentary film to better illustrate the circumstances, topics and content in a more motivated and appealing form.

Fig. 1: Screenshot from the documentary 'Underground City: London' that integrates 3D visualization elements

These new ways of presenting visual and dramaturgical content on TV can help viewers to better understand complex or hidden systems and promote interest in certain topics. Conventional linear Television remains a time based 'fluid medium' nevertheless. The advent of numerous recording or on-demand techniques makes it possible to pause, repeat and search for further additional related information pertaining to the real-time broadcast.

As stated above, based on the broad range of offered content and significant interest in the information genre, it is evident that the demand for 'informal learning' with TV based content will continue to grow.

More effective and more stable information acquisition and thereby improvements in the informal learning process will occur if 'viewsers' are not only watching, playing and stopping 'mono-medial' TV content, but are also able to access information via different medial formats and from different content sources, actively involving them across the entire spectrum of the knowledge acquisition process. Such 'effectiveness' aspects associated with the successful knowledge acquisition and construction are well documented in the spheres of e-learning and multimedia didactics. [7]

At this point provides Hybrid-TV in particular with the HbbTV technique the possibility for the audience to interact 'directly' with the viewing content to a) get additional information in variety presentation mode or from different sources or/and b) to use interactive application for deeper elaboration of the content and the presented problem/issue.

From a media-didactics perspective, however, it is well known that neither deep interactivity nor professionally prepared complementary multimedia content 'alone' are adequate in terms of ensuring effective information acquisition and knowledge construction, nor for an effective learning process. Particularly when 'informal learning' is applied to understand a wide range of different new topics in the TV 'information world' the audience

[1] Classifying audiences as 'Digital Natives', 'Digital Immigrants' or 'Millennium Children' is not uncontroversial, but is appropriate for the purpose of this paper.

needs to be provided with an overview of the specific issue. In the otherwise chiefly self-regulated constructivistic learning environments, the 'cognitive apprenticeship' approach is a major instructional & learning methods. The principle here is initially introduce, demonstrate, explain, help and guide learners to internalize problem-solving methodologies, and then to encourage them to act independently of the instructor in a step-wise fashion. Self-guided information and knowledge acquisition is particularly suited for learners familiar with the issue at hand or those requiring additional information. [2]

Other constructivistic well-known approaches include 'Anchored Instruction' and 'Problem based learning', which focuses on promoting an interesting issue at the start of the learning process in order to achieve deeper interest and motivation among the audience for the presented issue. Here, Television is very well suited as an 'anchoring and motivating engine' with its enormous imagery and dramaturgic potential.

Television broadcasts can be perceived according to the above mentioned approaches and aspects such as 'Motivation, Identification and Overview Streams' or 'Guided Tours' (spoken in 'old fashioned' Hypermedia language) with respect to a topic, an issue or a problem. With the technical options currently offered by Hybrid-TV systems like HbbTV and especially through the segment/topic related synchronized additional content or/and 2nd Screen application pertaining to the on-screen broadcast, 'viewsers' can immediately access additional information or become interactive participants via Apps, whether on the TV or on the 2nd Screen. Armed with these technologies, a wide range of effective informal learning methods such as those mentioned above can be offered in a meaningful mix of passive (e.g. introduction, overview, problematization) and active (e.g. additional multimedia content, involving apps) phases and artifacts of the converged TV and Online media environment.

In the two following projects below we have tried to create such mixed and converged interactive and multimedia TV/Online/2nd Screen environments for more effective informal learning.

PROJECT 1: AUGMENTED NEWS MAGAZINE THROUGH SEGMENT-RELATED & SYNCHRONIZED ADDITIONAL CONTENT & 2ND SCREEN

In this project – in association with the second German national public television broadcaster ZDF – we have developed a Connected Hybrid-TV & 2nd Screen real scenario based on the HbbTV standard for the 30-minute late-evening daily news magazine 'heute-journal'. The goal of the project was to give audience the opportunity to access additional more comprehensive, issue-synchronized information on the topic at hand during the live news broadcast (magazine or documentary). The topics generally run between two, and not longer than five minutes. The broadcast stream automatically triggers the TV and the 2nd Screen device to display additional information – initially as thumbnails or text lines – using 'stream events' techniques, which are integrated into the DVB transport stream.

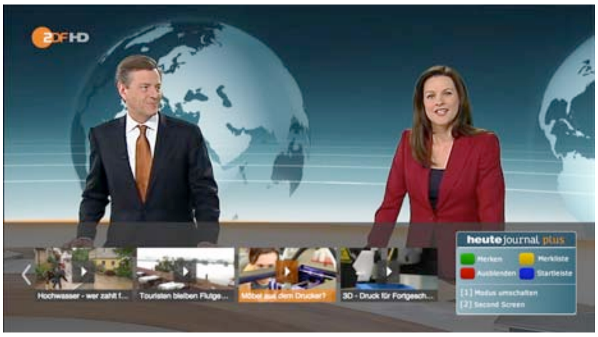


Fig. 2: Screenshot Project 1 'Connected HbbTV' Application with related content shown on the TV screen

The additional related content appears as a thumbnail tail in the lower area of the screen. A 'short mode' also provided that can display the supplementary incoming information in the non-intrusive form of a text line above the navigation panel (see the next figure).

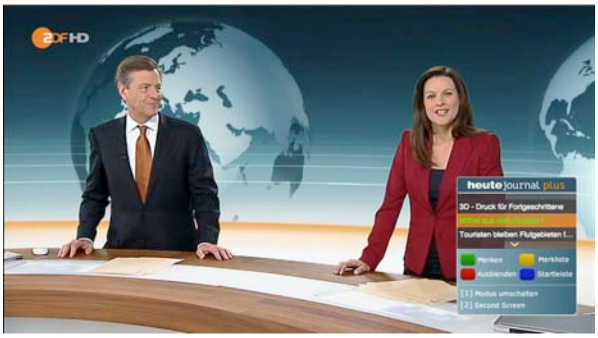


Fig. 3: Screenshot Project 1 in the 'short mode'

The 'short mode' should take up as little space as possible on the TV screen.

The related additional content will be prepared within an editorial process. ZDF is currently working to automate the content matching process.

In the 2nd Screen App (see figure 4 below) a time line corresponding to the news broadcast time is displayed, which will shifts automatically during the broadcast, as long as the user does not touch the screen. Synchronous to the stream events, fresh information appears at the corresponding time stamp on the running timeline.

In this way viewers still have very easy and intuitive access to the additional information concerning the current presented topic of a broadcast.

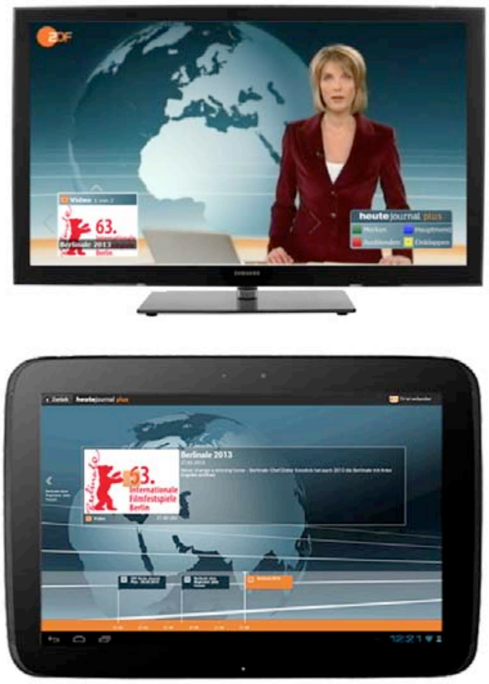


Fig. 4: The synchronized functionality of the 1st (above) and the 2nd Screen – on both screens same new content

Initially, the audience is not required to search at any website or media database for the additional information, thereby loosing focus on the running broadcast. Both the mobile and TV App present teasers of information related to the current running topic. The viewer can either access the supplemental information immediately via one-finger touch or by selecting the teaser with the RC or by bookmarking this information for time-shifted access to the information after the news broadcast.

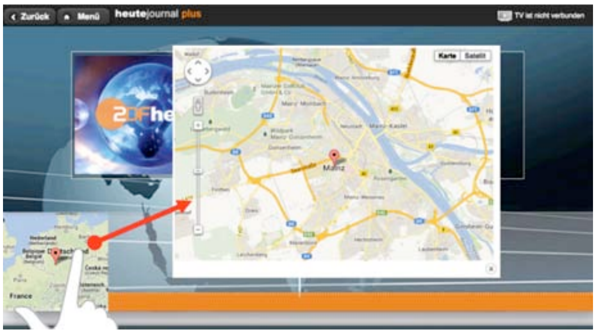


Fig. 5: Screenshot of the 2nd Screen App with the presentation of geographic information for each topic in the news

The viewer can decide to view the additional content on the TV Screen only, on the 2nd screen only, or on both screens. This helps to keep the TV screen as 'clean' as possible, particularly in when more than one viewer is watching the TV screen. The second advantage of the mobile presentation is the intuitive user interface that reacts to touch and swipe gestures to navigate the timeline in order to access the desired topic point.

The connection and the synchronization between the TV and the mobile device will be established through a QR code on the TV screen in the form of session number, which is scanned via the Tablet's camera. The communication then runs via a synchronisation server based on node.JS technology.

This News Application is an example of a 'gentle mix' of the passive and active components of a news magazine, indeed a non-disruptive and very flexible instrument in terms of providing and using supplemental content related to the viewed TV program.

At the present time we are conducting usability tests with the 2nd screen App.

PROJECT 2: SYNCHRONIZED AND CONVERGED 2ND SCREEN APPLICATION FOR INFOTAINMENT GENRE

The second project focuses on HbbTV 'application split' between a TV and a Tablet device to facilitate comprehensive and creative interactivity, especially for children.

The broadcast video shows a graffiti artist and his work; for this purpose we have created a 'gamification' and creativity App for 'graffiti'/painting tasks that functions on tablet devices.

The purpose of the application was to actively involve children in the topic of graffiti paintings and to give them the opportunity to create their own electronic graffiti paintings on several objects, for instance on an underground train wagon or a wall.

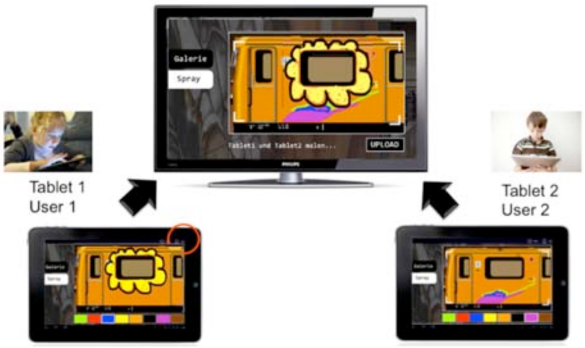


Fig. 6: Schematic visualization of the functionality of the project 2

Our Application promotes collaborative work, for example two or more children can work on one graffiti painting together and make the results available for all viewers of the application or the broadcast on demand.

The application belongs to both the 'Split' and 'Converged Apps' categories. 'Split' because one part of the application – the choice of painting object/background – is only possible on the TV device. The painting 'canvas' is only possible on the Tablet device. The gallery of user generated graffiti pictures can be run on both devices.

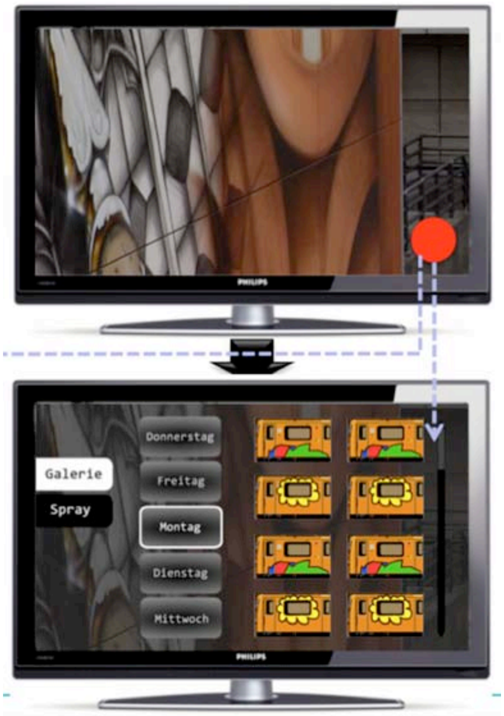


Fig. 7: During the presentation of the 'Graffiti Artist Film' the viewer can start the Graffiti-Application with the Red Button

The application is also 'Converged' due to a seamless connection between the TV and the 2nd Screen components of the application. Moreover, the 2nd Screen App is an extension of the 1st Screen App and is visual completely matched with the 1st Screen App.

In this project we have established a quite new converged technique between the 1st and the 2nd Screen, the so called 'Application Triggering' Effect.

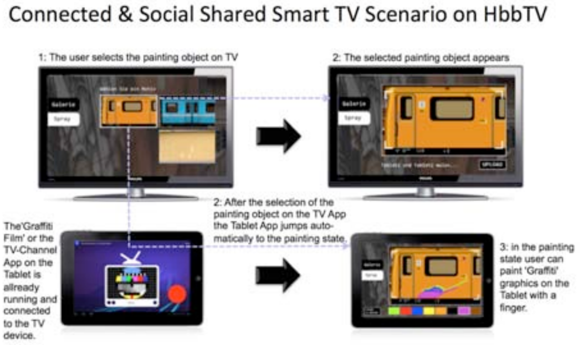

Connected & Social Shared Smart TV Scenario on HbbTV
1: The user selects the painting object on TV
2: The selected painting object appears
Galerie
Spray
2: After the selection of the painting object on the TV App the Tablet App jumps automatically to the painting state.
The'Graffiti Film' or the TV-Channel App on the Tablet is allready running and connected to the TV device.
3: in the painting state user can paint 'Graffiti' graphics on the Tablet with a finger.

Fig. 8: Schematic illustration of the functionality of the 'Triggering Effect' to change the 2nd Screen App from the 1st

This means that the 2nd Screen App can be triggered either by a certain time point during the broadcast and in the DVB / Transport Stream or through a special interaction/activity inside the 1st Screen part of the converged application. In our case we use the latter solution and trigger the 2nd Screen App to change the state of the application from the beginning / overview state to the graffiti/painting state, after the user has chosen on the TV screen the painting object. The chosen object appears on the TV screen but only in the presentation mode to show later the creative changes on this object. On the Tablet appears the object in the ready to interact canvas and in painting mode (see the figure above).

The painted object can be published to the joint gallery and made accessible for the application users, whether passive or active. The application has been tested during on two occasions and the children very excited about using this interactive infotainment offer.

CONCLUSION

An increasing number of technical options are available that can serve to promote information and educational content, in addition to novel media and interactivity solutions for modern informal learning TV & Internet based environments. Numerous options exist for the highly targeted synchronizing and linking of presented TV content (segment based) with supplemental information in the form of Apps for both TV and 2nd Screens. TV format developers, broadcasters and producers should therefore consider the great options at their disposal in terms of attracting the attention of 'viewsers'!

REFERENCES

[1] 2nd Screen Society (2014). The 2nd Screen: Transforming video consum. URL: http://www.2ndscreensociety.com/wp-content/uploads/2013/10/Q4-2013-Update-the-2nd-Screen-Transforming-Video-Consumption.pdf

[2] Duffy, T.M. & Cunningham, D.J. (1996). Constructivism: Implications for the design and delivery of instruction. In: Jonassen, D.H., Handbook of Research for Educational Communications-Technology, (pp..170-198). New York: Simon & Schuster MacMillan.

[3] Gerhards, M., Klingler, W., Blödorn, S. (2013). Sparten- und Formattrends im deutschen Fernsehen. In: Media Perspektiven, 4/2013, p. 202-220

[4] Gesellschaft für Unterhaltungs- und Kommunikations-elektronik gfu (2013). Deutschland vernetzt sich. URL: www.gfu.de (press area)

[5] Hölig, S., Hasebrink, U. (2013). Nachrichtennutzung in konvergierenden Medienumgebungen. In: Media Perspektiven, 11/2013, p. 522-536

[6] Marshall, C. (2013). Online Video Consumption Continues To Grow: Mobile, Tablet Use Booms [Report]. URL: http://www.reelseo.com/online-video-grows-q2-2013

[7] Mayer, E. (2009). Multimedia learning. Cambridge University Press

[8] nielsen (2012). Double Vision - Global Trends in Tablet and Smartphone Use while Watching TV. URL: http://www.nielsen. com/us/en/newswire/2012/double-vision-global-trends-in-tablet-and-smartphone-use-while-watching-tv.html

[9] Oehmichen, E. & Schröter, C. (2007): Zur typologischen Struktur medienübergreifender Nutzungsmuster. In: Media Perspektiven, 8/2007, p. 406 – 421

[10] SevenOne Media (2012). Navigator Mediennutzung 2012. URL: https://www.sevenonemedia.de/web/sevenone/research_ mediennutzung_navigator-mediennutzung

Vconect - Orchestration for Group Videoconferencing

Wolfgang Weiss
Institute of Information and
Communication Technologies
JOANNEUM RESEARCH
Graz, Austria
wolfgang.weiss@joanneum.at

Rene Kaiser
Institute of Information and
Communication Technologies
JOANNEUM RESEARCH
Graz, Austria
rene.kaiser@joanneum.at

Manolis Falelakis
Department of Computing
Goldsmiths, University of
London
London, UK
m.falelakis@gold.ac.uk

ABSTRACT

Current videoconferencing systems show a lack of support in adapting visual remote camera presentation to the users' needs. Some manage to put focus on the current speaker. In this demonstration we show an automatic decision making component in the realm of social video communication that aims to go beyond that. Our approach takes into account several aspects such as the current conversational situation, conversational metrics of the past and device capabilities to make decisions on the visual representation of available video streams. This allows to optimally support users in communication within various communication contexts.

Author Keywords

Videoconferencing, communication orchestration, automatic decision making

INTRODUCTION

Audio-visual communication pervades slowly but continuously our daily life, mainly driven by the availability of broadband services and mobile devices. One important aspect in our life is social communication with our friends. This type of communication is featured by certain characteristics for example people could join and leave continuously, the dynamic of the conversation might change over time and the network capabilities might change. Current video communication systems for social communication have limited intelligence to adapt to specific communication situations. We argue that taking into account conversational metrics and other parameters such as device capabilities and to adapt the visual representation of the videoconferencing client accordingly helps the user to get immersed by the communication experience.

The Vconect[1] project investigates novel ways of supporting mediated audio-visual communication for ad-hoc groups. One problem implied by such video communication setups is that for each participant, there are multiple video streams available as options for being currently shown, i.e. when there are n participants and each is equipped with 1 camera, $n-1$ exterior video streams are candidates for being displayed at each client (no self-view assumed). The question is how to optimally deal with them. An intuitive, but not always scalable option, is to show each user all video streams side by side on one screen (referred to as tiled layout). We set out to investigate more sophisticated solutions with the aim of achieving better communication support through intelligent camera selection. When taking into account further parameters such as conversational metrics [1] which represent the dynamics of a conversation, and the capabilities and size of the client it is possible to select a suitable visual layout and the corresponding video streams. A component which automatically executes a mixing process of different video streams is known as a Virtual Director [2]. In the realm of communication this is mostly referred to as Orchestration [4].

Subsequently we discuss the architecture and influencing parameters of the decision making component which intelligently switches between visual representations and executes the video mixing process for each user. Then the demo is described in detail and highlights what users can expect.

ORCHESTRATION

Communication Orchestration is the decision making which controls the mixing process of all available video streams. It can be compared to compiling a live TV transmission but in the case of video conferencing it has to address communication rather than narrative needs. Orchestration is a reasoning process which operates in real-time for each location participating in the communication. It builds upon the audio and video processing infrastructure and executes camera control and audio-visual composition (cf. [4]).

The Orchestration Engine in the Vconect project automatically produces camera selections and visual layout changes by reasoning on audio-visual cue streams from its participants. The system is implemented as a three step process:

Cue extraction: Audio-visual streams are processed by analysis modules in the system underneath and low level cues are extracted in real-time. An example for a low level cue is "voice activity".

Fusion and interpretation: Low-level cues from all locations are aggregated in the Semantic Lifting module of the Orchestration Engine. In this stage, higher-level semantic events that concern the communication as a whole, such as a "turn-shift", are generated while properties of the state of the interaction at each point, such as "active speaker"

[1] http://www.vconect-project.eu/

2nd International Workshop on Interactive Content Consumption at *TVX14*, June 25, 2014, Newcastle, UK.

or “crosstalk”, are evaluated continuously. The module aims to achieve a computational interpretation of the current communication situation on a semantic level that can directly be evaluated by the decision making components. The Semantic Lifting module also calculates conversation metrics such as “turn shifts per active participants” or the “active participation duration per each participant” based on a sliding temporal window. These conversation metrics allow to identify monologue situations or to identify the “heatedness” of a discussion.

Decision making: The application of mixing rules that result in the shot selection and the selection of the visual layout is made by the Director modules. For each screen one separate instance reasons based on high-level events and conversation metrics received from Semantic Lifting and other modules. This process allows to select the optimal visual layout in combination with the necessary video streams for each user and gives best support in communication by respecting the given limitations.

The two latter components are part of the Orchestration Engine which is a central, server-side software component. The logic is implemented declaratively using forward-chaining rules of the JBoss Drools[2] reasoning engine. Detailed examples of rules already incorporated into the system together with challenges about their implementation were reported in [3].

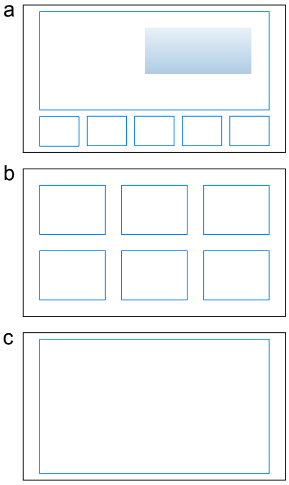


Figure 1. (a) Layout with focus on one person and with smaller tiles for all other participants. (b) Tiled view layout. (c) Full screen layout.

THE DEMO

This demo shows the abilities of a video conferencing system for social communication that takes into account various system parameters as well as conversation metrics to select the optimal visual layout for each participant. Figure 1a illustrates a layout with focus on one person and with smaller tiles for all other participants. This is suitable for most situations when the screen size is big enough and the number of participants does not exceed more then 10 people and the conversation is in a low or normal pace. If the conversation gets more heated (animated), meaning there is a high number of turn shifts between the participants within the analysis time window, it would be beneficial for the user to see all involved participants on the screen. This visual layout is illustrated in figure 1b but this can only be applied if certain other parameters allow a switch e.g. there needs to be enough space on the users screen. A single full screen layout which has only one video stream (see figure 1c) will be chosen when there is a monologue detected by a participant. Another reason for a single full screen layout would be if the the screen size is very small.

Interested persons have the possibility at the live demo to use one videoconferencing node on site to join a video conferencing session together with four remote participants. The remote participants are located in different offices in different countries. Video streams are transmitted in HD quality from all participants. Users can experience a videoconferencing system which automatically selects a suitable visual layout and the right video streams to optimally support users in communication. It is also possible to manually select the visual layout so that the automatic decision making system can be easily compared.

ACKNOWLEDGMENTS

The research leading to these results has received funding from the European Community’s Seventh Framework Programme (FP7/2007-2013) under grant agreement no. ICT-2011-287760. We thank all partners who contributed to the Vconect system for their support to make this demo possible.

REFERENCES

1. Hammer, F., Reichl, P., and Raake, A. The well-tempered conversation: interactivity, delay and perceptual voip quality. In *Communications, 2005. ICC 2005. 2005 IEEE International Conference on*, vol. 1 (May 2005), 244–249 Vol. 1.

2. Kaiser, R., and Weiss, W. *Media Production, Delivery and Interaction for Platform Independent Systems: Format-Agnostic Media*. Wiley, 2014, ch. Virtual Director.

3. Kaiser, R., Weiss, W., Falelakis, M., Michalakopoulos, S., and Ursu, M. A rule-based virtual director enhancing group communication. In *Multimedia and Expo Workshops (ICMEW), 2012 IEEE International Conference on* (July 2012), 187–192.

4. Ursu, M. F., Groen, M., Falelakis, M., Frantzis, M., Zsombori, V., and Kaiser, R. Orchestration: Tv-like mixing grammars applied to video-communication for social groups. In *Proceedings of the 21st ACM International Conference on Multimedia*, MM ’13, ACM (New York, NY, USA, 2013), 333–342.

[2] https://www.jboss.org/drools/

Using the SIVA Suite as a Multimedia Help System for Technical Applications in SME

Britta Meixner, Christian Handschigl
Chair of Distributed
Information Systems
University of Passau, Germany
meixner@fim.uni-passau.de,
admin@handschigl.com

Stefan John
Professorship for Media
Computer Science
University of Passau, Germany
stefan.john@uni-passau.de

Harald Kosch
Chair of Distributed
Information Systems
University of Passau, Germany
harald.kosch@uni-passau.de

ABSTRACT

In this demo paper we present how the SIVA Suite can be used as a multimedia help system for technical applications in SMEs. After describing our use case, a mechanics scenario, we show how our software was extended to fit all requirements of this scenario. We present short overviews over each component of the SIVA Suite: the authoring tool, the player, and the server application. Thereby, important new features are described briefly.

Author Keywords
Multimedia; Help System; HTML5 Player; Hypervideo; Server Application

ACM Classification Keywords
H.5.2 Information Interfaces and Presentation: User Interfaces—*Interaction styles*; H.5.4 Information Interfaces and Presentation: Hypertext/Hypermedia—*Navigation, User issues*

INTRODUCTION

Dynamic presentations like videos are well suited for the explanation of procedural skills, especially motor skills [1]. Most of the traditional videos show the execution of a procedural task in one single video. If subtasks are already known by the viewer, finding the right subsequence with the needed information can be a laborious task. Tasks consisting of multiple subtasks are hard to navigate due to the lack of an overlying structure like a table of contents or a search function. Tasks where single steps are depending on certain conditions result in duplicate scenes in different videos. This results in high download volumes using online videos, because usually large parts of a video are downloaded until the scene with the needed information is found. These problems can be overcome with hypervideos in combination with navigational elements. Longer videos are split up into scenes and a navigational structure as well as additional information to video contents is added. In the proposed demonstration, we show how our software can help finding information for the repair of a desktop computer more easily and quickly.

RELATED WORK

Several tools exist, which are capable of producing and playing hypervideos, but they are either not implemented with recent technologies or lack certain needed features for our scenario. Two important tools from related work are Hyper-Hitchcock [5] and Klynt [2]. Hyper-Hitchcock is a desktop tool for detail-on-demand videos. Here, an explanatory video is shown and the viewer is able to retrieve further information on a certain task. An additional video is loaded on request, at the end of which the main video is resumed. Klynt is capable of producing graph-based links between videos and adding different types of annotations to a currently shown video. The player is implemented in HTML5, which allows the playback of the videos on a large number of different end user devices. Klynt is only available under a proprietary license and cannot be extended and adapted to the requirements of our computer repair scenario or other scenarios.

SIVA SUITE

The SIVA Suite consists of three parts which will all be presented in the demonstration. New features of all parts, compared to previous work, are described briefly in this section. Multimedia instructions are created in the authoring tool and uploaded to the server application. The player can work in two modes, it either downloads the instruction and works in offline mode, or it downloads a control file and required multimedia files when they are needed.

SIVA Producer

The authoring tool called SIVA Producer was improved in different areas compared to the version presented in [4]. The video framework was updated and missing features were implemented. The settings dialog as well as the export dialogue were simplified. The text editor was replaced by our own implementation. The menu bar was extended with additional functions like a graph checker and the color layout of the whole application was unified. The editor for the markings in the video which display an annotation after a user click was revised as well (see Figure 1).

2nd International Workshop on Interactive Content Consumption at TVX14, June 25, 2014, Newcastle, UK.

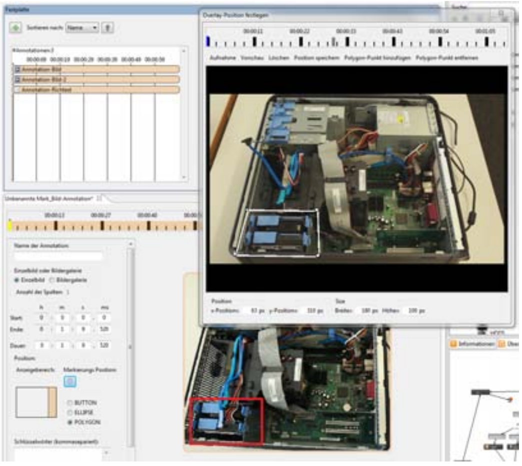
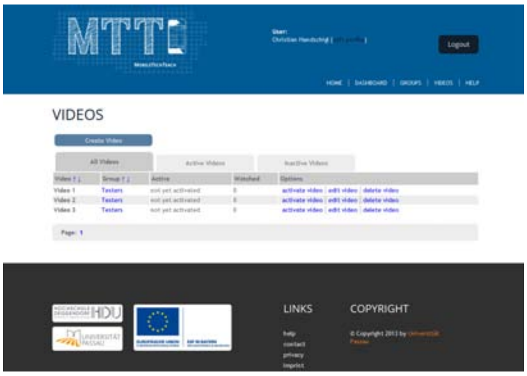


Figure 1. The annotation editor for the marking annotation.

Server Application

The server application is used to manage users, user groups, and videos (see Figure 2). A rights management is implemented to ensure that the visibility of videos is satisfied according to the demands of the author. Certain materials have to be protected from unauthorized access to protect the copyright. The server furthermore provides the backend for the logging functionality as well as interfaces to export the logged data in different formats.



VIDEOS

Video ↑↓	Group ↑↓	Active	Watched	Options
Video 1	Testers	not yet activated	0	activate video | edit video | delete video
Video 2	Testers	not yet activated	0	activate video | edit video | delete video
Video 3	Testers	not yet activated	0	activate video | edit video | delete video

Page: 1

Figure 2. The video overview in the server application.

SIVA Player

With the implementation of HTML5 in most browsers, even those for tablets and smartphones, it is now possible to implement one player and use it for all platforms. This makes the maintenance and updates of our previously used players implemented in Flash [4] or for Android devices [3] unnecessary. Our new HTML5 player has a simplified layout with one main annotation area on the right side and one navigation area on the left side. Both can be hidden or shown separately. Central buttons are grouped according to their range of applicability. Those needed in a scene are positioned at the bottom, those for the whole video at the top of the player (see Figure 3). A logging functionality is implemented to track the user behavior for individual videos. The logged data are synchronized with the server. We use JSON for the definition of the hypervideo control file.

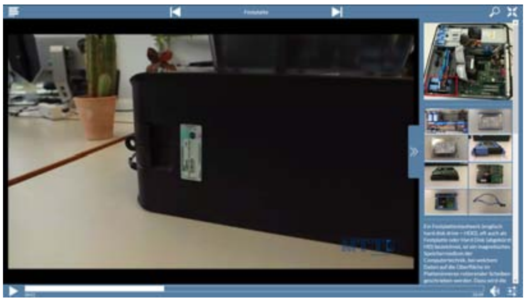


Figure 3. The player with the annotation area on the right side and the extended button bars at the top and at the bottom.

DEMONSTRATION

In this demonstration, we show how the SIVA Suite can be used as a multimedia help system for a technical application. We use a computer repair scenario for our illustrations. We present the authoring tool, the new HTML5 player, and the server application with focus on new functions compared to previous work.

ACKNOWLEDGMENTS

This work was partially supported by European Social Fonds and the Bayrisches Staatsministerium für Wissenschaft, Forschung und Kunst (Bavarian State Ministry of Sciences, Research and the Arts) under project name “MobileTechTeach - mobile multimediale Hilfesysteme für technische Anwendungen in KMU” (“MobileTechTeach - Mobile multimedia help systems for technical applications in SME”).

REFERENCES

1. Ayres, P., Marcus, N., Chan, C., and Qian, N. Learning hand manipulative tasks: When instructional animations are superior to equivalent static representations. *Computers in Human Behavior 25*, 2 (2009), 348 – 353. Including the Special Issue: State of the Art Research into Cognitive Load Theory.

2. Honkytonk Films. Klynt. Website (accessed April 2, 2014). http://www.klynt.net/.

3. Meixner, B., Köstler, J., and Kosch, H. A mobile player for interactive non-linear video. In *Proceedings of the 19th ACM International Conference on Multimedia*, MM ’11, ACM (New York, NY, USA, 2011), 779–780.

4. Meixner, B., Siegel, B., Hölbling, G., Lehner, F., and Kosch, H. Siva suite: Authoring system and player for interactive non-linear videos. In *Proceedings of the International Conference on Multimedia*, MM ’10, ACM (New York, NY, USA, 2010), 1563–1566.

5. Shipman, F., Girgensohn, A., and Wilcox, L. Authoring, viewing, and generating hypervideo: An overview of hyper-hitchcock. *ACM Trans. Multimedia Comput. Commun. Appl. 5*, 2 (Nov. 2008), 15:1–15:19.

Social Documentary: An interactive and evolutive installation to explore crowd-sourced media content

Fabien Grisard
UMONS
Mons, Belgium
fabien.grisard@umons.ac.be

Ceren Kayalar
Sabanci University
Istanbul, Turkey
ckayalar@sabanciuniv.edu

Sema Alaçam
Istanbul Technical University
Istanbul, Turkey
alacams@itu.edu.tr

Özgün Balaban
Singapore University of
Design and Technology
ozgunbalaban@gmail.com

Yekta İpek
Istanbul Technical University
Istanbul, Turkey
yektaipek@gmail.com

Stéphane Dupont
UMONS
Mons, Belgium
stephane.dupont@umons.ac.be

ABSTRACT

This paper aims to present a project in progress, an interactive installation for collaborative manipulation of multimedia content. The proposed setup consists in a vertical main screen and a horizontal second screen, which is used as control panel, reproducing an augmented physical desktop. Augmented reality markers are used to give the user an intentional way to interact with the system and a depth camera is used to estimate the users' gaze and quantify how interested they are in the displayed content, slightly modifying the video projection itself.

Author Keywords

Interactive installation; Fiducials; Gaze tracking; Attention estimation

INTRODUCTION

With the development and the diffusion of new technologies, many events (social, politics, sport, etc.) involve the production of huge amount of multimedia content. Thus, using them as raw material in artistic works is becoming more challenging but more interesting. The installation we propose is a way to display and to interact with big amount of content, mixing videos from different sources, images and texts. The visitor acts in two ways on the content : by choosing explicitly notions related to the video to display and in a much more unconscious way, by being (or not) interested in it.

RELATED WORKS

Augmented reality marker, also called fiducials, lead to the rapid development of new possibilities in table-top augmented tangible user interfaces. The web is full of examples of use of tangible interfaces in several domains such as artistic and musical creation[1] or educative games[2]. Usually, the artistic installations don't deal with the visitor attention. We want to use it along with the fiducials, as an input of our system. A fiducial is a geometric unique two-dimensional figure which provides presence, orientation, location and identity information in real-time, when placed in the field of view of a camera.

INSTALLATION DESCRIPTION

When the visitors enter the installation, they see a big screen (main screen) on the wall and a table on which a map is projected. There are also three sets of colored objects near the table, corresponding to three classes of keywords: people, action, emotion. Figure 1 presents an illustration of the installation with the detailed components. The visitors select one object of each color and put them on the map. A video segment related to the chosen keywords is displayed on the main screen. When the visitors are looking at the same area, a text is added to the video. Each time a visitor shows interest for a segment of the video, the rating of this segment is incremented, increasing the probability to display it to the next visitors. When all the visitors are gone, the main screen is shut down.

Our videos, images and texts are related to the social and political events that happened at Gezi Park in 2013. During this period, multimedia content have been produced in abundance both by the press and by demonstrators themselves, as mentioned by Alaçam et al.[1]. Videos taken during these events can provide historical documentaries putting together subjective and different views of the same event. Compared to classical historian-made documentaries, this novel approach provides emotion, hope and excitement captured during important moments which make the History.

Video segmentation and annotation

The video collection we found on Facebook mixes images from many sources. Each file is about three-hours-long. We chose not to cut them into short parts but to attach a subtitle-like file which contains useful information about each segment (beginning and end times, viewers interest rate - as Facebook "likes", annotation keywords). Thanks to the European project LinkedTV[3] visual analysis techniques [4], a big part of this work can be automatized. Those techniques include automatic shot segmentation and concepts recognition. The resulting file need to be manually edited to perfectly fit

[1] http://modular-drops.tumblr.com/

[2] http://www.woutersontwerpers.nl/portfolio/bosinfocentrum-t-leen/

[3] http://www.linkedtv.eu/

2nd International Workshop on Interactive Content Consumption at TVX14, June 25, 2014, Newcastle, UK. Copyright is held by the author/owner(s).

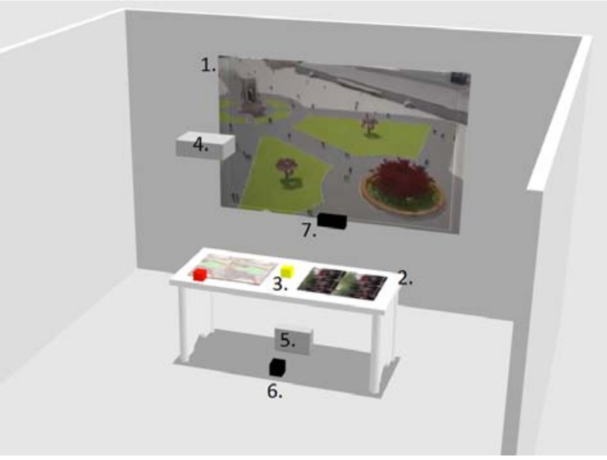


Figure 1. Installation setup with: 1. Main screen, displays the video stream; 2. Second screen, displays a map / Control panel; 3. Augmented reality markers on colored objects; 4. Main video projector; 5. Second projector; 6. webcam for fiducials tracking; 7. MS Kinect sensor for visitors' head tracking

our purpose but, thanks to the subtitle-like format, this part is quite easily done with any subtitle editor. By default, all the segments' rates are set to one.

Attention tracker

As exposed in the introduction, we propose an interactive installation based both on fiducial markers and on the analysis of the visitor attention for the currently played multimedia content. According to Hawkins [2], the time the visitor spends looking at the screen is linked to the attention he has for the projected content. Hawkins defines four "Types of Looks" : *Monitoring (¡= 1.5 seconds), Orienting (1.5-5.5 sec.), Engaged (5.5-15 sec.) and Stares (¿15 sec.)*. To identify which part of the installation the visitor is roughly looking at, we rely on the estimation of his head pose. Accordingly to Murphy-Chutorian [3], "[...] *Head pose estimation is intrinsically linked with visual gaze estimation* [...]. *By itself, head pose provides a coarse indication of the gaze that can be estimated in situations when the eyes of a person are not visible* [...]." The configuration of the installation (the screen width is close to the visitor-screen distance), increases this effect. As a result, if we can estimate the head position and orientation, we can infer on interest of the visitor. To achieve this goal, we use a Kinect sensor and the MS Kinect SDK face tracking functions[4]. Each time a visitor's attention is engaged (duration ¿ 5.5 seconds), the rate of the corresponding segment is incremented.

Segment selection and effects

Any object from wooden or plastic material can be turned into a tangible controller by sticking a fiducial under it. The application displaying the video content is a client listening to the TUIO (UDP based network protocol)[5] messages from the reacTIVision framework[6]. The video content on the vertical projection changes according to the markers put on the table. When the number of objects on the table changes, a list of video segments is updated and contains all the segments annotated with the keywords corresponding to the objects. To avoid rapid convergence toward a reduced list of "popular" segments, the one we display is chosen semi-randomly, accordingly to its rate. Once the segment is chosen, the player jump to the beginning time of this segment.

In case of joint attention (two visitors looking at the same screen), we propose to display related tweets on the concerned screen and to increase the sound when the video shows protesters, as an intensification of the "voice of people".

If the user puts more than one item on the table, our system can relate these items together by the distance between them, give a visual feedback on the table and display a combined content on the video projection.

CONCLUSIONS

This project aims to provide a tangible desktop application to understand, display and interact with a big collection of multimedia content. The installation should be able to adapt itself to the public thanks to attention evaluation and an automatic video rating system.
In our case, the content is related to very localized social events but the system could be used for other purposes, as in museums. This system brings new opportunities to build subjective crowd-based documentaries.

ACKNOWLEDGMENTS

This work is supported by the Integrated Project LinkedTV (www.linkedTV.eu) funded by the European Commission through the 7th Framework Program (FP7-287911).

REFERENCES

1. Alaçam, S., Ipek, Y., Balaban, Ö., and Kayalar, C. Organising crowd-sourced media content via a tangible desktop application. In *MMM (2)*, C. Gurrin, F. Hopfgartner, W. Hürst, H. D. Johansen, H. Lee, and N. E. O'Connor, Eds., vol. 8326 of *Lecture Notes in Computer Science*, Springer (2014), 1–10.
2. Hawkins, R. P., Pingree, S., Hitchon, J., Radler, B., Gorham, B. W., Kahlor, L., Gilligan, E., Serlin, R. C., Schmidt, T., Kannaovakun, P., and Kolbeins, G. H. What produces television attention and attention style? *Human Communication Research 31*, 1 (Jan. 2005), 162–187.
3. Murphy-Chutorian, E., and Trivedi, M. M. Head pose estimation in computer vision: A survey. *IEEE Trans. Pattern Anal. Mach. Intell. 31*, 4 (Apr. 2009), 607–626.
4. Stein, D., Apostolidis, E., Mezaris, V., Abreu, N. D., and Mller, J. Semiautomatic video analysis for linking television to the web. In *In Proc. FutureTV Workshop* (2012), 1–8.

[4] http://msdn.microsoft.com/en-us/library/jj130970.aspx
[5] http://www.tuio.org/
[6] http://reactivision.sourceforge.net/

I Remember/Know/Guess What I Saw: A False ‘Belief’ Technique to Features Selection

Jangsun Lee
Research Institute of Serious Entertainment (RISE),
Department of Industrial Engineering
Hanyang University.
222 Wangsimni Ave.
Sungdong-Gu, Seoul,
Rep.Korea
jaylee14@hanyang.ac.kr

Jieun Kim
Graduate School of Innovation and Technology Management,
Hanyang University.
222 Wangsimni Ave.
Sungdong-Gu, Seoul,
Rep.Korea
jkim2@hanyang.ac.kr

Hokyoung Ryu
Research Institute of Serious Entertainment (RISE),
Department of Industrial Engineering
Hanyang University.
222 Wangsimni Ave.
Sungdong-Gu, Seoul,
Rep.Korea
hryu@hanyang.ac.kr

ABSTRACT

In this position paper we address issues with the primary decision problem in the Smart TV UI design – feature selection. While the existing feature selection methods that traditionally make up HCI research were not able to render what features are to be prioritised in the new TV design, we will introduce the ‘*False belief technique*’ for this advancement. This new experimental technique will greatly enable UI/UX researchers to conduct feature selection evaluations that could effectively examine a users’ schema of the smart TVs, in a rather unconscious way at the expense of extra training time, which are unimaginable before.

Author Keywords

Smart TV; feature selection; schema; false memory; DRM paradigm; new product development

ACM Classification Keywords

H.5.2. Information interfaces and presentation (e.g., HCI): User interfaces---User-centered design

INTRODUCTION

A survey by Wiggin (Guardian Weekly issued at 24.July.2013) revealed that around 62% of British TV viewers are using “Twitter“ or “Facebook“ during TV watching. It is interesting to see that though many smart TVs already include such built-in functions for social chatting and networking, no more than 5% of the smart TV holders have ever used the functions, the report says. This lack of adoption of the smart features have much disappointed the smart TV designers, and they now seriously question of what features should be added or deleted for the utility of smart TVs.

However, no prior user-centered design for feature selection methods (such as interview, survey, card sorting) have fully suggested the smart TV designer of how to do this. The reason is, as Thompson et al [9] claimed, that we, as buyers, seem to overestimate how often we will use the overloaded features, and that we also underestimate how easily we, as users, will figure out how to use these functions in the future, believing that designers would make the functions not that difficult to use – an erroneous assumption. Such inconsistent attitude in user’s features adoption and selection has made such design decisions complicated when designers apply the current feature selection methods, so a practical technique to select appropriate features has been the key concern of the designers. The purview of position paper therefore is to introduce a practical features selection technique that implicitly utilize users’ newly formed schema, so that it can support the designer in what features should be added and further developed in novel product developments.

USERS’ SCHEMA AND FALSE MEMORY

In psychology and cognitive science, a schema can be described as a mental structure of preconceived ideas, a framework representing some aspect of the world, or a system of organizing and perceiving new information [1].

Memories are attributions that we make about our mental models based on our “subjective” qualities rather than “picture-like” qualities. Hence, they are often seen to be affected by our imaginative or intuitive beliefs, motives and goals, under a particular social context [4]. Bartlett [2] viewed ‘schemas’ as a major cause of this phenomenon.

This conception of schemas would be applicable to the first-time smart TV users. When firstly used, they would form a particular schema of the smart TV, mainly building upon a similar digital device experience or similar past events. If they do not have such existing experiences, they have to form an arbitrary schema to easily remember the features and how to use it. Thus, schema often frames

2nd International Workshop on Interactive Content Consumption at *TVX’14*, June 25, 2014, Newcastle, UK.

people to accept or reject new features or information, and serves as their own mental reference models.

An important note of false memory is thus further needed here. False memory can defined as remembering things that has not occurred or having a memory for an event which is distorted in some way [3]. Interesting is that the wrongly formed belief is more persistent over time than the correct memory unless it is fixed by repeated experiences or information [5, 8]. A series of the studies suggest that a *gist trace* that captures a thematic essence of the event decays more slowly, and claim that the false memory set in on the reinforcement of the thematic essence or meaning of the event or information [6]. This means that, when an event occurs, false belief in conjunction with gist memory is thought to establish a strong mental model of an event and it can be examined how this would be formed when one uses the smart TV.

THE *'FALSE BELIEF TECHNIQUE'*

The present paper addresses a practical technique to support the features selection technique using the 'false belief' theory. This is further built upon the '*Deese–Roediger–McDermott (DRM)*' paradigm [7]. The paradigm involves the visual presentation of a list of features in the target device, here, Smart TV and let the first-time users to watch tutorials of smart TV features, performed by an expert user. Like what DRM suggests, they were given a pen-and-paper recognition memory test a week later. In the recognition memory test, they were also asked to rate how confident they are about their answers as a 6 point Likert-scale. This DRM procedure would submit what the gist trace from a one-hour exploration is ready to construct the mental model of the smart TV.

During memory test, when more participants answer correctly on the 'listed' features (e.g., Internet surfing), having a full confidence with the answers and a higher mean confidence rating of the answers, one can consider the features are more likely to match with the schema the participants have formed.

In a similar vein, when more participants are hooked to the 'lured' (unlisted) features (e.g., photo sharing), having a full confidence with their answers and a higher mean confidence rating of the answers, they can be seen that the features are supported by the schema the participants have formed in a one-hour exploration of the smart TV.

Notable case is, if a lured feature is falsely recognized with high confidence rate at their recognition, that feature is called as a critical feature that should be put into smart TV, because it means that users could have chance to recollect the feature when they need to use it in the future.

DISCUSSION

The purview of this position paper is to suggest a practical feature selection technique that utilize users' newly formed schema in an unconscious way, so that it could support TV designers in what features should be added and further developed in future TV development.

We acknowledge that the process of *'False Belief Technique'* having been developed in our study is not certain nor the only way to induce participants to use their schema. Nevertheless, our proposition that designers should monitor how users build schema or related gist memories about their prior experience and should consider it for feature selection design is worth further developing. Opportunities and future developments in this area could be an interesting topic for discussion at the workshop.

REFERENCES

1. Anderson, J. R. *Cognitive psychology and its implications*. Macmillan, 2005.
2. Bartlett, F. F. C. and Bartlett, F. C. *Remembering: A study in experimental and social psychology*. Cambridge University Press, 1995.
3. Loftus, E. F. and Pickrell, J. E. The formation of false memories. *Psychiatric annals, 25,* 12 (1995), 720-725.
4. Mitchell, K. J. and Johnson, M. K. Source monitoring: Attributing mental experiences. *The Oxford handbook of memory*, (2000), 179-195.
5. Payne, D. G., Elie, C. J., Blackwell, J. M. and Neuschatz, J. S. Memory illusions: Recalling, recognizing, and recollecting events that never occurred. *Journal of Memory and Language, 35*, 2 (1996), 261-285.
6. Reyna, V. F. and Brainerd, C. J. Fuzzy-trace theory: An interim synthesis. *Learning and Individual Differences, 7*, 1 (1995), 1-75.
7. Roediger, H. L. and McDermott, K. B. Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 4 (1995), 803.
8. Seamon, J. G., Luo, C. R., Kopecky, J. J., Price, C. A., Rothschild, L., Fung, N. S. and Schwartz, M. A. Are false memories more difficult to forget than accurate memories? The effect of retention interval on recall and recognition. *Memory & Cognition, 30*, 7 (2002), 1054-1064.
9. Thompson, D. V., Hamilton, R. W. and Rust, R. T. Feature fatigue: When product capabilities become too much of a good thing. *Journal of Marketing Research*, (2005), 431-442.

Television Experience Insights from HbbTV

Jeroen Vanattenhoven
CUO Social Spaces
iMinds - KU Leuven
Leuven, Belgium
jeroen.vanattenhoven@soc.kuleuven.be

David Geerts
CUO Social Spaces
iMinds - KU Leuven
Leuven, Belgium
david.geerts@soc.kuleuven.be

Dirk De Grooff
CUO Social Spaces
iMinds - KU Leuven
Leuven, Belgium
dirk.degrooff@soc.kuleuven.be

ABSTRACT

In this position paper we shortly highlight the most important results from the European research project HBB-NEXT that concludes in March 2014: an ethnographically inspired user study provided many insights into the ways people use second screens; an experiment comparing gesture, voice, and remote control with Microsoft's Kinect showed that remote control still performs best, that voice looks very promising, and that gesture is useful in certain instances; an experiment with 51 participants validated four novel Social User Experience measures for the (user) evaluation of Group Recommender systems. Finally, we briefly outline our research agenda for another European research project, TV-RING, which started in September 2013. Both research projects focus on novel concepts and applications using the HbbTV Standard.

Author Keywords

HbbTV, User Experience, Multimodal Interaction, Group Recommendations, Second-Screen

ACM Classification Keywords

H.5.m. Information interfaces and presentation

INTRODUCTION

Traditional TV sets are currently being replaced by Smart TV's that use a variety of technologies and different Interaction paradigms. Game consoles act as media centers and will increasingly be able to deliver TV content. Second screens, desktop computers and mobile devices such as Smartphone's and tablets are ever more popular for watching on demand, streaming content. The question then is: How will we be able to provide usable systems, and provide positive user experiences to all users, given this diversity of systems and implementations?

In the HBB-NEXT project (http://www.hbb-next.eu), we tried to answer some of these questions. For starters, we conducted a thorough, ethnographically inspired user study into how people consume all kinds of content on any device in the home. Then, we conducted paper prototyping sessions concerning the general design of Smart TV interfaces, and specifically looked into group recommendations [2]. Finally, we also investigated different means of interacting with the Smart TV: gestures, speech, and the traditional remote control.

In the TV-RING project (http://tvring.eu), we are continuing along this path by working on contextual recommendations in the home, and ways to offer compelling second-screen content, again using the HbbTV standard. This project will carry out large pilots in the Netherlands, Germany and Spain, each one focusing on different aspect. Our work will align mostly with the Dutch pilot.

In the remainder of this position paper we will briefly sketch some of the important insights gained in the HBB-NEXT project. Afterwards, we will provide an outlook concerning the research agenda carried out in TV-RING.

DIARY STUDY

Our first research activity in the project concerned the gathering of broader user requirements for the to-be-developed applications in the project. For that purpose we carried out a three-week diary study in 12 households in Flanders [3]. The households comprised a mix of singles, couples, couples with very young children, and larger families with children. Participants were asked to report all their TV and video watching activities, the devices they use, with whom they were watching something, and possible second-screens that were used simultaneously. After the three-week diary period, we conducted interviews with all members of the household to discuss the diary entries. Interviews and diaries were then analyzed using quantitative analysis software (NVivo), and applying a Grounded Theory approach. After this analysis a number of main categories appeared.

In this study we will focus on a small set of results. We distinguished non-program related from program-related second-screen use. Mostly, second screen devices were used not in relation to the program on TV. Some people were working and kept half an eye on the program. Other people were on social networks conversing with friends and family. This relates somewhat to the next observation: people's attention switches quite frequently between the main screen and the second screen. Even within one

2nd International Workshop on Interactive Content Consumption at TVX'14, June 25, 2014, Newcastle, UK. Copyright is held by the author/owner(s).

program people will shift attention away from the TV screen when it is not interesting enough. Some participants explained that for the show "So You Can Think You Can Dance", they really enjoyed the dancing, but as soon as the program started looking behind the scenes, doing interviews etc. they lost interest and started doing other things. This emphasizes the importance of keeping your viewers engaged throughout the program. When creating second screen apps, it could be interesting to provide program-related content on the second screen that people can consume when losing interest in the program on TV to keep viewers engaged.

MULTIMODAL INTERACTION

For many years we have been using a remote control to interact with our television. In recent years however, other opportunities have been studied and implemented. The main methods of interaction here are speech and gesture control. For this purpose we conducted a comparison of speech, gesture and remote control, for a number of basic media task using Microsoft's Kinect together with the Xbox: navigating to a movie and play, searching for a movie and play, navigating to a song and play, and fast-forward 30 minutes into a video and play [1]. We measured task times, evaluated positive and negative elements of the interaction via a post-questionnaire with 30 users, and observed the interaction to identify usability issues with each interaction modality. Finally, after they carried out the tasks with all techniques, we asked them to provide a preference ranking.

As expected remote control performed best across the board. This had been our default way of using a TV for so many years; people really are very familiar with it. Yet, seven out of 30 people indicated that voice was their first choice. Voice control indeed performed really well. Especially for searching content, it is much better suited than either remote control or gestures. Furthermore, via voice control users have immediate access to the entire screen; using a remote control users first have to navigate left and right to get to the right item. Gestures did not perform so well. It was great for navigating to the next and previous screen using swipe gesture. The biggest downside was that people experienced fatigue up to the point some had to support their gesture arm with their other arm.

GROUP RECOMMENDER UX

Another important focus was the design of recommender systems. For many years the focus was very much on improving the accuracy of these systems. In recent years though, the focus has shifted toward improving their user experiences. For a group recommender then, it makes sense to evaluate the social user experience. Since no measures for Social UX were available, we introduced and validated four new measures in an experiment with 51 users: Social Choice Difficulty (how difficult was it for the group to agree on a choice), Anticipated Social Experience (to what extent do people expect a social experience from watching the chosen item), Togetherness (to what extent do people feel together), and Social Perceived Usefulness (how useful do people consider a group recommender for these purposes) [4].

TV-RING

In TV-Ring we are concerned with providing the right recommendations at the right time in the home. In order to achieve that we aim to map when each person's personal device is in the living room. That way we hope to determine who is in front of the TV, so we can automatically adjust the content to it. Related efforts will focus on the right user interface for such personalized, dynamic sets of recommendations. In addition, we hope to determine the appropriate time to offer recommendations – viewers enjoying their weekly show, do not have any use for recommendations at this time.

ACKNOWLEDGMENTS

The research leading to these results was carried out in the HBB-NEXT project and has received funding from the EC Seventh Framework Programme (FP7/2007-2013) under Grant Agreement n°287848, and in the TV-Ring project (EC grant agreement ICT PSP-325209).

REFERENCES

1. Müller et al. (2012) D6.2 Implementation of Mock-Ups and Early Application Design. Public Report. http://www.hbb-next.eu/index.php/documents
2. Vanattenhoven, J., Geerts, D., & De Grooff, D. (z.d.). Deciding What to Watch: Paper Prototyping Interactive Group Recommenders for Television. In Proceedings of TVUX-2013: Workshop on Exploring and Enhancing the User Experience for TV. Paris, France: ACM.
3. Vanattenhoven, J., & Geerts, D. (2012). Second-Screen Use in the Home: An Ethnographic Study. In Proceedings 3rd International Workshop on Future Television, EuroITV 2012 (p. 12). Berlin: Springer.
4. Vanattenhoven, J. (2013). D2.3.2 Annex 2 User Validation Results / Annex 2: Group Recommender. Public Report. http://www.hbb-next.eu/index.php/documents

Full Research Papers

Using SIVA XML and SMIL for Interactive Non-linear Videos: a Comparison

Emanuel Berndl
Chair of Distributed
Information Systems,
University of Passau, Germany
berndle@fim.uni-passau.de

Britta Meixner
Chair of Distributed
Information Systems,
University of Passau, Germany
meixner@fim.uni-passau.de

Harald Kosch
Chair of Distributed
Information Systems,
University of Passau, Germany
harald.kosch@uni-passau.de

ABSTRACT

With recent technologies, it is possible to create appealing multimedia presentations or extended videos with a high level of interactivity. Standards like SMIL provide extensive structures to describe metadata for timing and spacing of single media elements which then form a presentation. While multimedia presentations are viewed mainly in a linear manner, provide interactive and non-linear videos a much higher level of interactivity and navigational possibilities. In this work, we examine the expressiveness of SMIL for the support of interactive non-linear videos. It has to describe temporal and spatial relationships of videos and annotations, as well as interaction and navigational elements. We therefore compared SMIL with the SIVA XML. We tried to find ways to express SIVA XML structures with SMIL attributes and elements. After that, we compared the DTD/XSD of SMIL and SIVA XML using XML metrics. We thereby focus on the language implementations. We do not take their implementations in authoring tools or players into account. Concluding that SMIL has little disadvantages in terms of feasibility for interactive videos, we propose minor additions that could resolve these problems and make SMIL more appropriate for our use case.

Author Keywords

SMIL, Interactive Video, Non-linear Video, XML, Video Annotations, Multimedia Document, Metrics

ACM Classification Keywords

I.7 Document and Text Processing: Document Preparation—*Hypertext/hypermedia, Multi/mixed media, Standards*

INTRODUCTION

Nearly each web page provides multimedia contents today. These reach from animated images to sounds and embedded videos. With recent technologies and increasing Internet bandwidths, appealing combinations of different types of media and various forms of user interaction are possible. However, the contents of web pages are not temporally synchronized. Therefore, more advanced languages with spatial and temporal models are necessary. Two main fields of research can be found in this area, namely “multimedia presentations” and “hypervideos”. An applicable definition of multimedia presentation is given by Nimmagadda et al. as follows: “*Multimedia presentations are collections of different media files [...] like text, images, videos, and animations with different resolutions, durations, and start-times. [...] The layout of multimedia presentations is defined by the locations and the start times of the objects*” [16]. In contrast, interactive non-linear videos are defined by us (extended from [9]) as follows: “*[...] An interactive non-linear video is a digitally enriched form of video materials arranged for an overall concept. It presents additional information beyond the original content. Furthermore, it offers new forms of influence and navigation in the video and additional contents*” [15]. They are a subset of hypervideos for which many different definitions and descriptions can be found. A summarizing definition can be given as: Hypervideo is defined as video based hypermedia that combines non-linear video structuring and dynamic information presentations. Video information is linked with different kinds of additional information (like texts, pictures, audio files, or further videos). Users can mouse-click on sensitive regions (having spatial and temporal characteristics) within the videos to access the additional information (heterogeneous hypervideo) or jump to other scenes (homogeneous hypervideo). Hyperlinks build a graph between main video scenes and additional information.

Watching multimedia presentations, the viewer is rather passive, but basic interaction and navigation may be possible. The viewer is elicited from his passivity viewing interactive non-linear videos in contrast. This form of video consists of video scenes (“main video”) and additional information which enhance the scenes. Timeline and control bar are extended with additional functions. These provide control on the flow of the video and give hints on when additional information (in the remainder of this work referred to as annotations) is displayed. Decision elements in the video allow the selection of a certain branch of the video instead of watching it in a linear way. Furthermore, additional information which may be any type of medium, like text, image, or video, is added as an annotation to the main video. We proposed an XML format for this type of video in [13]. Its structure was designed for the definition of interactive non-linear videos while using SMIL [22] may be possible up to a certain point, but leads to problems and work-arounds in some areas. In this work, we try to show the advantages and disadvantages of using SMIL for the description of interactive non-

2nd International Workshop on Interactive Content Consumption at TVX14, June 25, 2014, Newcastle, UK. Copyright is held by the author/owner(s).

linear videos compared to our XML format[1]. An overview of authoring tools and players for SMIL can be found in our previous work [14]. We furthermore tested the only working player (Ambulant Player[2]), which showed weaknesses in the display of presentations as well as stability. Thereby, this work makes the following research contributions: Requirements were identified (see section REQUIREMENTS) and both formats were checked for their suitability to implement these requirements (see section FEASIBILITY ANALYSIS). Metrics were used to compare the complexity of the SIVA XML schema and SMIL, see section COMPARISON/METRICS.

REQUIREMENTS

An analysis of usage scenarios (further described in [12] and [14]) like e-learning, virtual tours, mobile help systems, or sport events revealed several requirements according to timing and spacing of media elements in interactive non-linear videos with additional information. Needed functions and elements are as follows (see also [13]):

- *Media, main video, and annotations*: As specified for interactive non-linear videos, usually one video is displayed as the main video. Additional information may be shown with this video. Therefore, several different types of media like images, audio-files, videos, and text should be usable. It should be possible to handle them differently during playback, for example should a subtitle be positioned automatically.

- *Event-based timing model*: Main video and annotations may be time dependent or time independent. For this reason, an event-based timing model is preferred to a structured timing model due to the high level of interactivity mixed with fixed points in time were annotations are displayed or hidden. By keeping timing issues as local as possible, synchronization is realizable more easily.

- *Temporal relationships between main video and annotations*: Temporal relationships in form of start and end point or durations of display need to be defined between the main video (scene) and each of the annotations.

- *Spatial relationships between videos and annotations*: A positioning of main video and single annotations or groups of annotations needs to be defined. Annotations may be displayed statically in areas around the video or as an overlay over the video. Furthermore, dynamic annotations may move on a path on the video canvas. Automated arrangement of annotations in defined areas facilitates the authoring process.

- *Decision elements at forks in video flow*: The playback of interactive non-linear videos includes different strands of scenes. Selection elements are needed to select the next scenes which are displayed to the viewer. Selection elements may be buttons or links.

- *Table of contents*: One way of extended navigation is provided by a table of contents which has to be defined and linked with single scenes.

- *Keyword reference list*: A second way of extended navigation is implemented with a keyword search. Keywords need to be linked with scenes or annotations in order to find information more quickly.

- *Extensibility*: The structure of the XML format has to be extensible in case of new ways of interaction that should be mapped into the model. Furthermore, changes in the XML file should be kept as local as possible in the structure without changing bigger parts of the existing file. Scripting is not considered as useful with respect to the affordance of an easy to use authoring tool.

DESCRIPTION OF THE FORMATS

The SIVA XML schema and the SMIL DTD show several differences in structure and scope. While SMIL tries to cover many different areas of application, the SIVA XML schema is exactly tailored to the needs of interactive non-linear videos with additional information. We now give a short overview over the formats before we compare them based on the requirements we determined in the previous section.

SMIL DTD

SMIL stands for Synchronized Multimedia Integration Language and it is a standard for interactive multimedia presentations released by the World Wide Web Consortium (W3C). Design goals of SMIL were to define “an XML-based language that allows authors to write interactive multimedia presentations. Using SMIL 3.0, an author may describe the temporal behavior of a multimedia presentation, associate hyperlinks with media objects and describe the layout of the presentation on a screen. [Furthermore, it should allow] reusing of SMIL 3.0 syntax and semantics in other XML-based languages, in particular those who need to represent timing and synchronization” [22]. Used media files are images, text, audio files, videos, animation, and textstreams which are linked to an internal graph structure. Navigation is possible in a presentation but not in single continuous media files. Furthermore, it is possible to define hotspots for navigation or to display additional information. With the usage of the elements and attributes from the timing modules, “time can be integrated into any XML language” [6, p. 117]. It is possible to define start and end time, duration, persistence, and repetition of objects and relations between those objects [6, p. 117]. The layout of a presentation is defined by the “relative placement of (multiple) media objects”, but SMIL does not concern the internal formatting media of objects [6, p. 149]. SMIL is based on CMIF [5] and the AHM [10]. The final version of this standard is the SMIL 3.0 Recommendation which was published on December 01, 2008 [22]. Previous versions of this standard were SMIL 1.0 released in 1998, SMIL 2.0 released in 2001, and SMIL 2.1 released in 2005 [6]. SMIL 3.0 consists of 12 modules of elements and attributes (Animation, Metainformation, Content Control, Structure, Layout, Timing and Synchronization, Linking, Time Manipulations, Media Objects, Transition Effects, smilState and smilText) described as a DTD [6]. Furthermore, five profiles are

[1] For a more detailed description of the SIVA XML schema see [13], SMIL is described in [6] and [22].

built which use the enlisted elements and attributes, namely the SMIL 3.0 Language Profile, the SMIL 3.0 Unified Mobile Profile, the SMIL 3.0 DAISY Profile, the SMIL 3.0 Tiny Profile, and the SMIL 3.0 smilText Profile [22]. These profiles may limit the elements and attributes of the standard or extend it with functionality from other XML languages [6].

Extensions for SMIL can be found in the work of Cazenave et al. [8], Pihkala and Vuorimaa [17], and Vaisenberg et al. [21]. These works add a table of contents, a search function, and a bookmark function [21], "location information, tactile output, forms, telephoning, and scripting" [17], and the option to publish multimedia documents on the web using HTML5, CSS, and SMIL Timesheets [8] to SMIL. We do not consider these language extensions, because they are not part of the standard. In the following sections only elements and attributes from the SMIL 3.0 specification are used.

SIVA XML Schema

The SIVA XSD[2] was designed during the projects "Interaktives Video Editierungstool zum netzwerkbasierten Wissenstransfer (ivi-Pro)"[3] and "iVi-Pro 2.0 - Interaktives Video im Zeitalter von Mobilität und Kollaboration"[45]. Major design goals were an easy expandability and a slim format which exactly fitted our requirements as well as existing and potential future scenarios without too many limitations. We decided to implement some logic into the player to avoid repetitive definitions in the XML file. This allows the XML files to be more flexible, easy to read and adaptable to the requirements. Besides a main video, usable media files are images, audio files, videos, rich texts, and subtitles. These can be displayed as "global annotations" during a whole video, or as "local annotations" during a single scene. It is possible to define a non-linear structure of scenes, whereat decision elements provide selection panels or quizzes to viewers. Other navigational elements are a table of contents and a keyword search. Hotspots in the video trigger the display of additional information. The timing is kept local by adding annotations to single scenes. Thereby, absolute times in the scenes are used for displaying and hiding annotations. Thus synchronization is only necessary for a single scene and not for a whole video. The SIVA XML consists of six parts represented by six main elements below the root element: `<projectInformation>`, `<sceneList>`, `<resources>`, `<actions>`, `<tableOfContents>`, and `<index>`. These elements are linked by `ID/IDREF` attributes which are checked by constraints for their consistency. A more detailed description of the XML format can be found in [13]. In contrast to SMIL, the SIVA XML schema is not an official standard.

Comparison of the SIVA XML Schema and SMIL

We are aware that in general, the SIVA XML schema is a more specific, focused, and limited approach, while SMIL is more general, flexible, and not by default made to fit our use cases of interactive non-linear videos. Both languages do not cover the same aspects, and have different focuses and levels of detail.

We do not take other models and formats like NCM/NCL [7, 19, 20], CHM [18], ZYX [4], and HTML5 [23] into account, because they are either not standardised or have other areas of application. SMIL in contrast is standardised and considered as a format capable of describing hypervideos in the multimedia community.

FEASIBILITY ANALYSIS

We already stated what an interactive non-linear video is and what requirements need to be fulfilled in order to satisfy the demands of such the videos described and analysed in the REQUIREMENTS section. The following part shows how feasible an implementation of an interactive non-linear video is with both of the given XML languages SMIL and SIVA XML. Therefore we first present the feasibility of every requirement in regard of both languages. We also propose how extensive the implementation is and what features can or can not be realised. Therefore we will especially emphasize on disadvantages in the specified requirements, which show the main points that disallow a fully satisfying solution for interactive videos. On the other hand, these points also show starting points for better adaptiveness towards this use case. Afterwards we conclude our feasibility results in the Feasibility Conclusion section. Examples used in this section are adapted from [3].

Media, Main video, and Annotations

The entire presentation of an interactive video consists of a main video with the addition of annotations. Annotations are multimedia elements, that are supposed to enhance the interactive feeling and can be used to give more information about the topic of the video. Annotations can be triggered (invoked for display) by user interaction, established by a click on certain defined portions of the video, or by reaching a specified point of time. The placement is a fixed point or a path, resulting in a moving annotation.

Both XML languages support the full variety of media annotations, but differences are met in terms of the placement. When the editing of a **SMIL** presentation is finished, the placement of all its elements is set. This can result in overlapping annotations, for example pictures, when they are placed in the same area. The **SIVA XML** is usually interpreted by a player which supports an automated placing function, that will arrange the annotations next to each other. Another weakness of SMIL is found concerning the pathing. In order to achieve the exact demanded movement, the element of the annotation would require four `<animate>` for each step of the path.

Event-based Timing Model

The **SIVA XML** is fully designed to fulfill the requested timing model that an interactive non-linear video needs. Scenes

[2] The XSD file can be downloaded from **http://siva.uni-passau.de/sites/default/files/downloads/sivaPlayer.xsd** (accessed May 12, 2014)

[3] "Interactive video editing tool for network-based knowledge transfer (iVi-Pro)"

[4] "iVi-Pro 2.0 - Interactive video in the age of mobility and collaboration"

[5] The development of the SIVA XML schema was partially funded by the European Social Fonds and the Bayrisches Staatsministerium für Wissenschaft, Forschung und Kunst (Bavarian State Ministry for Sciences, Research and the Arts)

are built modularly and do not have to be processed in a linear order as in SMIL. Annotations are started by defined triggers during a scene.

In contrast, **SMIL** makes use of an interval-based timing model. Although the whole functionality of an interactive non-linear video could be implemented, there are slight disadvantages with this model. Each relation between main video and annotation is bound together as a result of the SMIL element structure. It is not as modular and as local as in the SIVA XML.

Temporal Relationships between Main Video and Annotations

Temporal relationships can be implemented well in both languages. The modularity of the **SIVA XML** makes it possible that every temporal relationship can be modeled by local XML constructs which are linked by `ID/IDREF` attributes.

SMIL on the other hand supports a broad range of elements to satisfy the temporal needs of an interactive non-linear video. By making use of the basic temporal elements `<seq>` and `<par>` combined with more complex ones like `priorityClass` and their timing attributes `start` and `end`, each relationship between and inside the parts of an interactive video can be implemented.

Spatial Relationships between Main Video and Annotations

The **SIVA XML** shows advantages in spatial relationships compared to SMIL. All media and navigational elements of the interactive non-linear video can be placed specifically where they need to be, annotations can be arranged automatically, their paths and/or positions that are defined in the SIVA Producer will be fulfilled. If elements of a displayed panel (e.g. the table of contents) cannot be shown in its full size, the player can adapt to it by using techniques like scrollbars to supply the full range of accessibility for all elements.

SMIL on the contrary has some difficulties establishing these requirements. Each element has to be aligned exactly with its `left`-, `top`-, `right`-, `bottom`-, `height`-, and `width`-attributes in order to determine its position. If the given set, for example a list of links or buttons, is too large, it cannot be displayed entirely. The portion that is too large for the displaying area will be cut out. Furthermore, links can not be sized according to the width that their text needs, so consecutively you can activate a link by clicking into the "free" area that is residing to the right side of the link, what can also result in faulty activations of links. The counterpart here would be to size the buttons to a fixed length, but if a given text exceeds that boundary, the text will be cut off as well.

Decision Elements at Forks

Enhanced navigational features of the interactive non-linear video are forks. These are usually button panels and quizzes. At a button panel the viewer can pick one option that determines the continuation of the main video. In a quiz, a row of questions with multiple choice answers is posed. Each answer will give a certain amount of points. The achieved sum will determine the continuation after the quiz. For these functionalities, different elements like the button or answer panel are needed.

In the **SIVA XML schema**, all of this is supported entirely by defined complex types. The modular structure of the XML file makes the scenes accessible by triggers which are linked to the buttons of a choice panel. The quiz functionality specifies questions together with their answers. The correct answers are marked and obtainable points per question are set. Furthermore, point ranges for a whole quiz are defined for the selection of the following scene.

All of these features can be implemented in **SMIL**. Forks and their paths are supported by button panels, that suffer the spatial disadvantage already mentioned in the Spatial Relationships section. Although complex in the XML structure, the quiz functionality is also realizable. But a problem arises from the fact, that a path after a fork may be an edge to an already played scene. Jumps inside the SMIL file are in need of a link element. In order to implement this, the viewer is confronted with a panel that has to be clicked so that the video can continue at an earlier point. This results in a break of the flow of the video. For the common consumer of videos, this is an inconvenience. Problems like this are not present in the SIVA XML.

Table of Contents

The table of contents contributes to the non-linear character of the interactive video. When displayed (after clicking a button), a panel with links in a tree structure is presented to the user. By activating one of these links, the corresponding scene will be played. Here, the **SIVA XML** allows the addition of sub entries for each item of the panel. In **SMIL**, a table of contents is constituted by a list of clickable links, so for every entry, a link will be created and then arranged in a top down manner in the specified area. This element is also suffering the spatial problems mentioned in the Spatial Relationships section.

Keyword Reference List

A keyword search could not be established for the **SMIL** export as such functionality is not supported by the language. The **SIVA XML** supports all the requirements that are needed for a search. Users have the possibility to search for strings as keywords. Keywords are linked with scenes or annotations. When the user selects a keyword of a scene, the scene starts at its beginning. Selecting the keyword of an annotation, the video starts play-back at the point where the annotation is displayed. They can be searched while the interactive non-linear video is played.

Extensibility

Regarding the possibilities to extend the given model for new features, both XML languages are capable of integrating additional sets of elements. On both sides, the DTD or XSD files have to be adapted as well as the interpretation of the resulting exported XML document at the player. But in terms of the possibilities of changing an XML document of both languages that already exists, the **SIVA XML** has slight advantages. Due to the modular structure, it is easy to add new scenes, keywords, or annotations to an interactive video. This process is more complicated in **SMIL** based on the nested composition of the elements. Especially the temporal structure needs to be kept correct. Adding one single element may have impact on different parts of the interactive video, altering it in a way that may not be intended by an author or an insertion algorithm.

Feasibility Conclusion

Table 1 represents an overview of our analysis by listing each requirement and its feasibility for both languages. The feasibility is ranged on a scale from "*very bad or not at all*" (denoted as "- -"), "*partly feasible*" (denoted as "-") over "*neutral*" (denoted as "0") to "*feasible with some drawbacks*" (denoted as "+") or "*meets all requirements*" (denoted as "++").

Table 1. Feasibility of the Requirements

Requirement	SIVA	SMIL
Media, Main video, and Annotations	++	+
Event-based Timing Model	++	+
Temporal Relationships	++	++
Spatial Relationships	++	0
Decision Elements	++	+
Table of Contents	++	0
Keyword reference list	++	- -
Extensibility	++	++

The feasibility analysis shows, that the **SIVA XML** is very well adapted to the requirements of an interactive non-linear video. **SMIL** is able to realize many of the requirements or more precisely the needed features as well, but it lacks in certain details. As Table 1 demonstrates, in terms of temporal relationships or the extensibility, both languages are suited very well. The lack of a keyword feature sets SMIL back in that requirement, while facets like the problems in the spacial relationships force an inferior evaluation compared to the SIVA XML in other categories.

COMPARISON/METRICS

In order to have a closer look and a numerical comparison of the two metadata formats, we will make use of the following XML metrics: Size, Structure Complexity, Depth, Fan-In, and Fan-Out. For a detailed description see [11].

With these five values for an XML description format you are able to make statements about the complexity, comprehensibility, reusability, and convertibility of it. The higher one of the metric count is, the more complex is the possible resulting XML file. A high Fan-Out value makes it harder to alter a format because changes in single entities or elements may have an impact on multiple locations in the file. We have manually identified the count of the results for both, the SIVA XML and SMIL, to be able to compare them. For the first one, the SIVA XSD was converted into a DTD to be able to compare it with the SMIL DTD. Knowing that such a conversion will usually have an impact on the accurateness of the file, it does not have an impact on the evaluated metrics. The DTD for SMIL can be found online as well [1]. As it contains the elements for the whole language, we have created a profile to model a DTD that only supports the modules and elements needed in the export for interactive non-linear videos. SMIL was also contemplated in two different ways: with and without the `<metadata>` element, which has a great impact on the Fan-Out value. This is caused by the fact that the element can become a child node of each SMIL 3.0 element. The use of this element is the wrapping of structured meta information. It contains an own XML tree as content and is not processed by the player at all. As we do not make use of this element in our export, it is still contained in the generated DTD of our interactive video profile. Therefore we differentiate between an analysis of SMIL with and without the `<metadata>` element. Our results are presented in Table 2. Each number depicts the count for the specific XML metric.

Table 2. Comparison of SMIL and the SIVA XML (SC = Structure Complexity, ∞ = unbounded)

	Size	SC	Depth	Fan-In	Fan-Out
SIVA XML	58	67	5	12	8
SMIL (w/o meta)	40	430	∞	21	16
SMIL (with meta)	41	507	∞	22	38

Regarding the first two entries in Table 2, one can see that SMIL gets by with less elements than the SIVA XML, but its complexity is much higher. This is caused by the fact that many SMIL elements are used recursively. The high Fan-Out value is applicable for many of the occurring elements. The potential depth for SMIL is unbounded because the temporal elements `<par>` and `<seq>` can be boxed repeatedly. In some depth analyses, the recursion is ignored to not achieve a depth that is unbounded. We do not take this into consideration because in fact there can be an apparent endless potential depth by nesting forks. The Fan-In and Fan-Out metrics also state higher values in the SMIL-DTD and therefore indicate superior complexity.

PROPOSAL FOR EXTENSION

The extensibility of SMIL allows the addition of new elements that could be used to generate a module that is more adapted and able to provide support for interactive non-linear videos with regard to the aforementioned criteria. The ideas for these elements arose from the problems that were encountered while modeling the elements which are already available in the SIVA XML schema for interactive non-linear videos in SMIL. In combination with these elements, SMIL could achieve a more dynamic structure and be more feasible referring to interactive non-linear videos. Possible useful additions might be the following.

Jumps in the XML File

The elements `<goto>` and `<end>` may change the flow of the SMIL presentation by jumping to another position of the same document when reached. While the former cause a jump to a given `ID` supported by a `to` attribute, the `<end>` element is supposed to bring the presentation to an end. They can be used like most of the timing elements in SMIL in terms of nesting as well as their attributes. Caution has to be paid, because loops and abrupt endings of the presentation can be built very easily.

Choices at Forks

To satisfy the requirement for fork elements more easily, we propose the introduction of two new elements that could facilitate the implementation of a fork: `<fork>` and `<choice>`. When a `<fork>` element is started, it composes a standard choice panel (that could be altered by its attributes in terms of shape etc.) which contains buttons to start one of its `<choice>` children nodes. One of these `<choice>` elements contains elements that are supposed to be played once it is activated. In combination with the preceding elements `<jump>` or `<end>`, different continuations after a path can be established as well. If this is not supported, the `<choice>` as well as its surrounding `<fork>` element will be ended and the succeeding element will be started, just like in a normal SMIL presentation. Listing 1 shows an example for the use of `<fork>` and `<choice>` elements combined with the above mentioned jump elements `<goto>` and `<end>`. The fork ranges from line 4 to 23 and contains three different choices. By selecting one of them, the code inside the corresponding `<choice>` element will be started. The button panel itself is designed by the attributes of the `<fork>` element to show round buttons with a size of 20 pixels. The panel will be displayed for 30 seconds. If no path is chosen in that timespan, according to the `defaultPath` attribute, the defined default path (in this example *path1*) will be played. The second path choice from line 11 to 16 contains a `<goto>` element in line 12 inside a sequential container with the effect, that after the other elements defined in line 13, an automated jump from line 14 to line 2 (according to the given `to` attribute). The third path in lines 17 to 22 consists of a parallel node containing some code and then an `<end>` element in line 20, which also owns a `begin` attribute with the value `20s`. This structure of elements induces the behaviour, that no matter what the content of the `<par>` element is, the `<end>` element will be started after twenty seconds, causing the presentation to terminate.

```
1  <smil xmlns="http://www.w3.org/ns/SMIL">
2  <head>
3    <!-- Any SMIL head content -->
4  </head>
5  <body>
6    <seq xml:id="start">
7      <!-- Any SMIL content -->
8      <fork shape="circle" size="20"
9        region="main_region"
10       dur="30s" defaultPath="path1"
11       xml:id="fork">
12       <choice xml:id="path1" after="#fork">
13         <!-- Any SMIL content -->
14       </choice>
15       <choice xml:id="path2">
16         <seq>
17           <!-- Any SMIL content -->
18           <goto to="#start"/>
19         </seq>
20       </choice>
21       <choice xml:id="path3">
22         <par>
23           <!-- Any SMIL content -->
24           <end begin="20s"/>
25         </par>
26       </choice>
27     </fork>
28     <!-- Any SMIL content -->
29   </seq>
30 </body>
```

Listing 1. Body excerpt of a SMIL file showing sample code for a fork and jumping elements

CONCLUSION

In summary, the SIVA XML shows advantages regarding to the usefulness for interactive non-linear videos. As stated in section FEASIBILITY, all of the requirements that are needed to fully implement an interactive non-linear video are met by the language. SMIL can realize most of the functionality (keywords could not be established natively), but it lacks in some details like spatial problems of decision boxes, the placement of subtitles or moving annotations. A further benefit of the SIVA XML is revealed by the analysis of the underlying DTDs of both formats. SMIL is composed by a much more complex set of entities which makes the construction and the understanding of the resulting SMIL documents harder. The temporal elements in particular hamper the process of modifying a given SMIL presentation, which was stated in the Extensibility section. Many parallel, sequential, and conditional elements are stacked and interwoven, so that the addition of a simple annotation can be a very complex venture. These points state, that the SIVA XML is better suited for interactive non-linear videos. But an important detail to mention is that SMIL is not meant or designed particularly to support interactive non-linear videos. The research we did here was based on the standard SMIL 3.0 model.

A new set of extensions would make SMIL more usable for interactive non-linear videos. Therefore, SMIL could be extended by different complex structures like a decision fork, a textual link, and an element which allows to jump inside the SMIL file (without the activation of a link). In combination with these elements, SMIL could achieve a more dynamical structure and be more feasible referring to interactive non-linear videos.

The SIVA Suite[6] contains a production software (SIVA Producer) that enables the user to design and create interactive videos. As this work shows that SMIL is well suited in order to be used as metadata format for interactive non-linear videos, we implemented a SMIL exporter into the SIVA Producer. Now it is possible to export the designed interactive video into a SMIL presentation (with the limitations described in this work).

REFERENCES

1. SMIL 3.0 DTDs. Website, December 2008. http://www.w3.org/TR/smil/smil-DTD.html (accessed March 25, 2013).

2. Ambulant open smil player. Website, June 2010. http://www.ambulantplayer.org/ (accessed May 12, 2014).

3. Berndl, E. Siva-xml-schema vs. smil - strukturmapping und exportimplementierung. Mastersthesis, University

[6] The SIVA Suite can be downloaded from: **http://www.siva.uni-passau.de/?q=node/66** (accessed May 12, 2014)

of Passau, Chair of Distributed Information Systems, Passau, Germany, October 2012.

4. Boll, S., and Klas, W. Zyx - a multimedia document model for reuse and adaptation of multimedia content. *IEEE Transactions on Knowledge and Data Engineering 13*, 3 (2001), 361–382.

5. Bulterman, D., Rossum, G. V., and Liere, R. V. A structure for transportable, dynamic multimedia documents. In *In Proceedings of the Summer 1991 USENIX Conference* (1991), 137–155.

6. Bulterman, D., and Rutledge, L. *SMIL 3.0: Flexible Multimedia for Web, Mobile Devices and Daisy Talking Books*. X. media. publishing Series. Springer, 2008.

7. Casanova, M. A., Tucherman, L., Lima, M. J. D., Rangel Netto, J. L., Rodriquez, N., and Soares, L. F. G. The nested context model for hyperdocuments. In *Proceedings of the 3rd Annual ACM Conference on Hypertext. HYPERTEXT '91*, no. December, ACM (New York, NY, USA, 1991), 193–201.

8. Cazenave, F., Quint, V., and Roisin, C. Timesheets.js: when smil meets html5 and css3. In *Proceedings of the 11th ACM symposium on Document engineering*, DocEng '11, ACM (New York, NY, USA, 2011), 43–52.

9. Hammoud, R. *Interactive video: algorithms and technologies*. Signals and communication technology. Springer, 2006, ch. Introduction to interactive video, 3–25.

10. Hardman, L., Bulterman, D. C. A., and van Rossum, G. The amsterdam hypermedia model: adding time and context to the dexter model. *Commun. ACM 37*, 2 (Feb. 1994), 50–62.

11. Klettke, M., Schneider, L., and Heuer, A. Metrics for xml document collections. In *XML-Based Data Management and Multimedia Engineering - EDBT 2002 Workshops*, vol. 2490 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, 2002, 15–28.

12. Meixner, B., Hölbling, G., Stegmaier, F., Kosch, H., Lehner, F., Schmettow, M., and Siegel, B. Siva producer - a modular authoring system for interactive videos. In *Proceedings of I-Know, 9th International Conference on Knowledge Management and Knowledge Technologies* (Graz, Austria, 2009), 215–226.

13. Meixner, B., and Kosch, H. Interactive non-linear video: definition and xml structure. In *Proceedings of the 2012 ACM symposium on Document engineering*, DocEng '12, ACM (New York, NY, USA, 2012), 49–58.

14. Meixner, B., Matusik, K., Grill, C., and Kosch, H. Towards an easy to use authoring tool for interactive non-linear video. *Multimedia Tools and Applications* (2012), 1–26.

15. Meixner, B., Siegel, B., Hölbling, G., Lehner, F., and Kosch, H. Siva suite: authoring system and player for interactive non-linear videos. In *Proceedings of the international conference on Multimedia*, MM '10, ACM (New York, NY, USA, 2010), 1563–1566.

16. Nimmagadda, Y., Kumar, K., and Lu, Y.-H. Preference-based adaptation of multimedia presentations for different display sizes. In *IEEE International Conference on Multimedia and Expo, 2009. ICME 2009.* (2009), 978–981.

17. Pihkala, K., and Vuorimaa, P. Nine methods to extend smil for multimedia applications. *Multimedia Tools and Applications 28* (2006), 51–67.

18. Sadallah, M., Aubert, O., and Prié, Y. Component-based hypervideo model: high-level operational specification of hypervideos. In *Proceedings of the 11th ACM Symposium on Document Engineering. DocEng '11*, ACM (New York, NY, USA, 2011), 53–56.

19. Soares, L. F. G., and Rodrigues, R. F. Nested context model 3.0. part 1 - ncm core. 18, Laboratorio TeleMidia DI - PUC-Rio, Rio de Janeiro, 2005. Technical Report. Website (accessed April 26, 2014).

20. TeleMidia Lab - PUC-Rio. Ncl - nested context language. Standard/Website, 2011. http://www.ncl.org.br/en/inicio (accessed Mai 12, 2014).

21. Vaisenberg, R., Jain, R., and Mehrotra, S. Smpl, a specification based framework for the semantic structure, annotation and control of smil documents. In *11th IEEE International Symposium on Multimedia, 2009. ISM '09.* (2009), 533–539.

22. W3C. Synchronized multimedia. Website, March 2012. http://www.w3.org/AudioVideo/ (accessed Mai 12, 2014).

23. W3C. Html5 - a vocabulary and associated apis for html and xhtml (w3c candidate recommendation 6 august 2013). Website, 2013. http://www.w3.org/TR/html5/Overview.html (accessed Mai 12, 2014).

Concept mapping second screens to augment understanding of science on television

John Dowell
Computer Science,
University College London,
j.dowell@cs.ucl.ac.uk

Sylvain Malacria
Computer Science,
University College London,
sylvain@malacria.fr

Sean O'Halpin
BBC R&D
Central Lab,
sean.o'halpin@bbc.co.uk

ABSTRACT

Science programmes on television are the main source for public understanding of science; they also make some of the greatest demands of TV viewers. Companion second screens running on connected, handheld devices have the potential to augment viewers' understanding of science programmes, for example by providing additional explanations or additional information about the programme. We examine two prior exemplars of second screen designs that were explicitly created to support viewer understanding of programme content. They suggest primary representational functions that a second screen should provide if it is to augment understanding of science TV programmes. We describe an interactive concept map as a second screen that shows promise for achieving those representational functions.

Author Keywords

Companion second screen; public understanding of science; augment understanding; user interface adaptation; interactive concept maps.

ACM Classification Keywords

H.5.2. User interfaces: theory and methods.

INTRODUCTION

Television programmes are the primary means of increasing public understanding of science [6,1]. If they are to be effective in this role, they need to be engaging and interesting and they need to be understandable for a very broad audience.

Science programmes on TV also make some of the greatest demands of viewers. At one level and in purely cognitive terms, viewing a science programme involves interpreting a presented fact base through understanding a nexus of more or less explicit scientific concepts informed by the viewer's background knowledge. Much of the art of the producer lies in designing TV science programmes that accommodate the diverse limits of the audience in relation to these processes.

Science TV programmes are a rich potential application for companion second screens. Second screens, running on connected, handheld, personal devices provide their viewer users with synchronized content as well as enabling them to participate in online activities and share comments with other viewers. They are likely to become an accepted part of watching TV on a large screen, having already demonstrated their potential for making TV viewing more active and participative [4,8].

Of course producers have no control over who will watch their programme and different viewers will have very widely varying background knowledge. A family watching a popular science programme might include a child not understanding much of the programme because of his or her lack of basic background concepts, whilst a parent finds the programme to be largely uninformative because it doesn't engage with knowledge of the subject they already have. Simply put, children may want more explanation while adults may want more information. In addition, viewers have different approaches to the use of TV programmes as an authoritative source of knowledge about science [6].

Development of second screen techniques is being pursued particularly actively by commercial interests, for example, to push synchronized adverts and offers to viewers. To date, programme makers have used second screens mainly for sports programmes, game shows and quizzes, and reality TV shows [8,10]. Second screens have been demonstrated with fact-based programmes including natural history programmes, but so far as we are aware not for science programmes and not for augmenting the viewer's comprehension. This prospect raises the question of in what form and by what method could a second screen augment the viewer's understanding of a science programme?

We address this question by identifying primary representational functions of second screens for science programmes. These functions are assayed from a review of two second screens previously developed to support viewer understanding. The representational functions we extrapolate from these prior systems take account of the kind of cognitive processes involved in understanding a science programme on television.

TWO EXEMPLAR SECOND SCREENS

Two notable designs of second screens for augmenting understanding of programme content give insight into the

2nd International Workshop on Interactive Content Consumption at TVX'14, June 25, 2014, Newcastle, UK.

representational functions required of second screens for science programmes. The first exemplar is the second screen developed for an episode of Frozen Planet [1]. This app pushed content to the viewer in synchrony with the programme; with each animal introduced in the programme, a thumbnail image/ icon appeared in the app, giving access to a summary of facts about the animal which could be bookmarked.

Figure 1. Frozen Planet companion second screen [1].

Findings from the evaluation trial of the Frozen Planet second screen included the following [1]:

- it broadened and deepened viewers' engagement with the programme content;
- it allowed users a personalised viewing according to their interests;
- it encouraged curiosity and shared discovery;
- it provided additional but not excessive authoritative information;
- it worked as a log of items of interest to a particular viewer and marked things to access later;
- it particularly engaged 'pre-family' viewers (who sought more information) and children (who used it to continue learning journeys);
- it could compete negatively for the viewer's attention and was ignored during highly visual parts of the programme;
- smartphones rather than tablets were used to view the second screen during the programme, the converse for viewing the second screen later;
- simply repeating the broadcast content was not valued and additional content was expected;
- the second screen interaction should be user paced, not system paced.

The second exemplar is the Story-Map second screen for long form drama [11]. The app was designed to give viewers of serialized dramas an overview of the setting and narrative, to remind them of plot threads and to allow them to review important story sequences across episodes. Its main representation was an updating map of characters synchronized with the TV content to identify the people in the story and their relationships. As new characters appeared, they were added as a new node to a graph in which all the characters and their relationships were shown.

Characters who were mentioned but who had yet to appear were greyed. Characters who had appeared but were not currently present were indicated by a reduced size icon. Rather than showing who is literally visible in the frame, the map showed who was present in the semantic dramatic unit. Characters were also grouped according to geographical location and were iconified, providing access to a brief biography which was updated with the narrative. No evaluation trials with Story-Map have been reported.

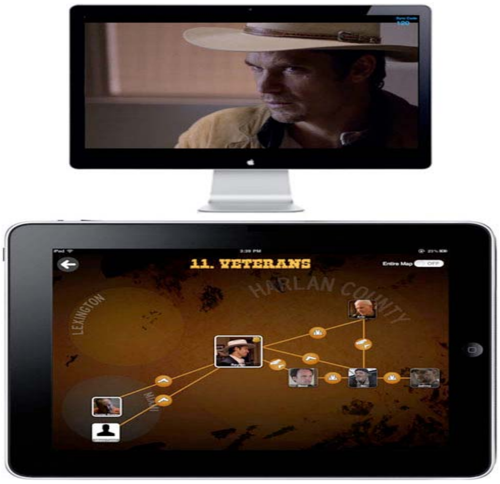


Figure 2. Story-Map's character map with the broadcast programme in the background [11]

Both the Frozen Planet and Story-Map systems were developed to support the viewer's understanding of the programme content. The Frozen Planet screen extended users' knowledge beyond the programme and assisted them in recalling what they had seen and learnt in the programme. By contrast, the Story-Map screen more intentionally provided an explicit representation of the content of the watched programme and provided explanation of that content supporting understanding of the story, in particular in explaining the relationships between characters. Both programmes involved understanding a set of facts (even if in one case those facts were about fictional characters and events). In this respect both programmes have similarities with science TV programmes. However they do not share the concepts and language that characterise science programmes and which are typically specialist, abstract and complex. We now identify a set of representational functions for second screens supporting science programmes that take account of these features,

REPRESENTATIONAL FUNCTIONS

Creating a second screen application to augment comprehension of a science programme is a challenging problem as the design space is large. By extrapolating from our review of exemplar systems above, we identify the primary representational functions a second screen for science programmes should provide.

Synchronized & coordinated

Synchronization is the most distinctive feature of second screens. Since second screens are used while watching a broadcast programme, their displayed content should be synchronized with the programme's current content. Input capabilities and user interactions should also benefit from this context of synchronization. For science programmes, it is of particular value to synchronize the second screen with the broadcast programme in order to (a) augment the concepts that are currently presented and (b) offer the viewer possibilities to interact with the second screen.

Both the Frozen Planet and StoryMap second screens were synchronized with the broadcast programme and were viewed in a passive and discretionary mode by the viewer user. The Frozen Planet study reported that user's access of the second screen should be self-paced rather than system-paced, implying that the screen updating should avoid being attention-seeking.

Coordination refers to the designed combination of contents of the screens to take account of the viewer's focus of attention. When not coordinated, users will develop their own ways of dividing and managing their attention between the screens, for example, by focusing on the second screen when the broadcast programme is showing only a presenter's 'talking head'. A coordinated second screen implies a requirement on the broadcast programme itself to provide opportunities for the viewer to look away. The Frozen Planet evaluation reported that viewers would ignore the second screen during highly visual sequences of the broadcast.

Persistent & accumulating

Viewers of science programmes need a view of the content they have already viewed. Television is, of course intrinsically transient and viewers' recall of what they have viewed needs to be augmented since science programmes often involve relating together their parts within a linear narrative. Therefore, the second screen needs to provide a persistent representation of the content already viewed that can be re-consulted. Moreover, the representation should be accumulating dynamically so that the content currently viewed extends the second screen representation. This dynamic leading edge would allow the viewer to easily find their place when shifting visual attention. This aide memoire function of second screens was provided by Story-Map; by contrast, viewers of Frozen Planet were found not to want only a repetition of the content they had seen broadcast, although that particular programme usually does not involve back reference to earlier content for understanding.

Synoptic

If, as we claim, viewers need a representation of the broadcast content already watched, a verbatim transcript and frame-by-frame record is unlikely to be helpful. Rather, viewers need an abstraction of the content that summarises the important ideas and their connections, and the significant facts in the broadcast that relate to these. The process of understanding the broadcast may be recognized as fundamentally constructing this conceptual abstraction over the content. We therefore identify the need for a synoptic representation of the programme. This representation would also provide a means of categorizing, indexing and therefore re-finding content already viewed.

The StoryMap second screen provided a synoptic representation of the drama in terms of the characters and their relationships and visually identified meta level features of the characters. The Frozen Planet second screen used the separation of the broadcast into sections dealing with different animals and events. Simple navigation through the second screen was provided by the thumbnails as a highest level synopsis.

Interactive (navigable, interrogable, indexable)

A second screen interface has the potential for making the usually passive viewing of a science programme into a active learning experience, engaging the user and improving overall satisfaction. In the context of science programmes, interactivity could offer the possibility to navigate and interrogate the flow of concepts and notions that have been introduced in the TV programme so far. Interactivity clearly has to be designed carefully as extended interactivity will lead to distraction that will undermine comprehension of the programme. Both the Frozen Planet and StoryMap systems enabled users to navigate through a representation of the programme content and to access additional information. Frozen Planet allowed users to add bookmark indexes to personalize the content for later referral.

Adaptable and adaptive

As already discussed, different viewers will have different needs and expectations for how the science programmes should be augmented. Viewers with a little knowledge would need more explanation to properly understand the programme, while viewers with better knowledge would need more information. For this reason the second screen should be adaptable so that the viewer can adapt its content to his or her expectations and needs. Ideally, the second screen should be able to infer these needs and automatically adapt. This adaptivity could be achieved by modeling the user using a combination of information about his or her background and monitoring the interaction.

CONCEPT MAPPING SECOND SCREENS

The design space of second screens for science programmes is clearly very large. Interactive concept maps are one possible kind of design, one we believe is capable of offering the desired representational functions.

A concept map is a form of directed graph used to capture and present knowledge in specific domains. Ideas and information are represented in abstract as labeled boxes or circles connected with labeled arrows, often in a downward-branching hierarchical structure. Relationships between concepts can be typed and annotated using constructs such as *is-a*, *has*, *causes*, etc. Using these constructs a concept map is able to model semantic associations between concepts and propositional knowledge (for example, the proposition 'an iMac is an apple' could be expressed with two nodes and one relationship).

Concept maps are related to, but distinct from, mind maps and argument maps [5]. Mind maps express ideas and the associations between ideas only loosely, sometimes using images as nodes rather than labels and not labeling the relationships; at the other end of the spectrum, argument maps often have well defined syntactic constraints which allow definite argument structures (e.g., claims, rebuttals, etc) to be expressed over sets of nodes and links.

Concept maps are a practical application of semantic networks [1] and have been advocated as a tool for learning in the sciences in particular [12]. They are probably most often used in the classroom in a constructional mode with students drawing maps of some learning contents they have been exposed to. In this mode they have also been used as a way of assessing knowledge and learning; stereotypical morphologies ('spoke', 'chain' and 'network' structures have even been identified [7]). But concept maps are not limited to a constructional mode of use; they can also be used in presentational mode, as a way of summarizing a set of taught material to aid learning.

Concept maps can equally be created for television science programmes and an example of one is shown in Figure 3. It represents the first three minutes of a programme on supernovae, an episode from a series on the lifecycle of stars [2]. The map was created using the concept mapping tool Cmap [3] and its nodes represent the main ideas explicitly stated in the narration overlaying a set of related video images in this opening part of the programme. The narration has been parsed and re-expressed into these separate nodes then relationships fixed between them. The figure uses colour coding to differentiate the nodes for each successive minute of the programme. The orange shaded nodes would therefore all appear in the first minute of watching the programme.

This map could be presented dynamically on a second screen so that nodes appear serially, synchronized with the play of the recorded programme. There are many possibilities for the visual transitions in displaying the map. For example, it might be presented as a fixed global view that is progressively filled in with blossoming nodes and arcs. Alternatively, the view might be a roving porthole that skims across the map revealing the network of nodes 'already there'. Both a roving porthole and dynamically appearing nodes could be used. Multiple detailed presentation design decisions arise, for example, the limiting speed of transition of the moving map without disrupting reading.

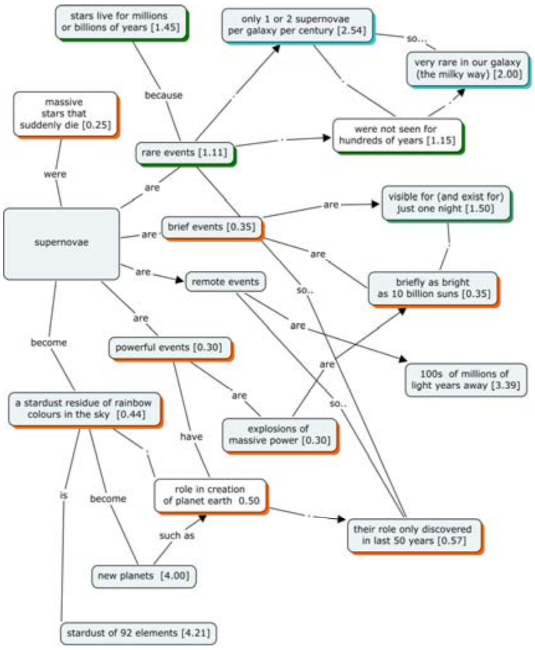

stars live for millions or billions of years [1.45]
only 1 or 2 supernovae per galaxy per century [2.54]
so...
very rare in our galaxy (the milky way) [2.00]
because
massive stars that suddenly die [0.25]
were not seen for hundreds of years [1.15]
rare events [1.11]
were
are
visible for (and exist for) just one night [1.50]
are
supernovae
are
brief events [0.35]
are
are
remote events
so..
briefly as bright as 10 billion suns [0.35]
are
are
become
powerful events [0.30]
are
100s of millions of light years away [3.39]
are
a stardust residue of rainbow colours in the sky [0.44]
so..
have
explosions of massive power [0.30]
is
become
role in creation of planet earth 0.50
such as
their role only discovered in last 50 years [0.57]
new planets [4.00]
stardust of 92 elements [4.21]

Figure 3. Concept map of the first three minutes of *Supernovae,* Episode 5 of *The Seven Ages of Starlight [2]*.

This concept map can also be interactive. At a minimum, viewers will be able to pause the dynamic display of the map. The map would then need to catch up with the programme. Beyond this simple touch-and-hold interaction, viewers could navigate the map and expand and mark nodes. Interactive concept maps have been previously demonstrated as a knowledge elicitation tool [9] but the second screen application greatly expands the possibilities for interaction.

We can also anticipate how the concept mapping second screen can achieve the desired representational functions identified earlier. Its synchronization with the science programme has been described, with nodes and branches coming into view at the right moment. Coordination would require the TV programme to be created with use of the concept map taken into account, and cued for example by a visual flag or by reducing the attentional demands of the programme. A viewer pausing the concept map could also be a signal to the programme to pause.

The concept map second screen provides a persistent representation of the programme content already seen and an accumulating representation through its controlled exposure. The map necessarily abstracts the rich audio-visual programme into a synoptic representation. The possibility for drilling down on a node to reveal more detail means that it is a synopsis of varying abstraction.

Navigation, interrogation and indexing of the concept map are possible during viewing of the programme and of course separately after viewing the programme. Familiar methods for navigating geographical maps and hypertext can be provided, as can common methods for browsing and searching visual databases. The value of simple indexing of items, marking favourite items etc, has already been proven for second screens. The possibility to *star* a topic of interest could be used to improve the relevance of additional content offered after the programme has been watched.

Adaptation in the second screen to respond to the viewer's understanding of the programme is also feasible. By providing choices for viewers in navigating and browsing the map, the second screen is offering a basic form of adaptability. If the choices available to the viewer are constrained by their individual profile, or by the trajectory of choices already made through the map, then the map is achieving at least a rudimentary degree of adaptivity.

Many open questions remain about the form of concept mapping second screens. If done badly it is entirely possible that a concept mapping second screen would be ignored, or worse, make comprehension of the TV programme harder. But if the primary representational functions are provided, we believe the concept map will be effective. A multitude of detailed interface design issues are apparent. For example, it is currently uncertain whether the relationships between nodes should be of only a few stock kinds, or not typed at all. It is even unclear if links should have an annotation at all for presentation on a second screen.

CONCLUSION AND FUTURE WORK

Television is the most important channel we have for supporting the public's understanding of science. Although TV science programmes are abundant, their production is a matter of compromise in order to reach the largest audience. As a result, science programmes can be both too easy for some viewers and too challenging for others.

A companion second screen offers the ideal interface for augmenting the science programme, supporting a more active engagement, augmenting viewers' grasp of the programme and extending their knowledge beyond the broadcast content. To date however, use of second screens for science programmes has not been explored.

In this paper, we examined two prior second screen designs intended to augment viewer understanding. Although the programmes they accompanied, one a natural history and the other a crime drama, may not offer the same intellectual challenge as science programmes, they gave insight into five representational functions we believe a second screen for science programmes should. We then proposed the concept map as a particular form of second screen display for science programmes, one we argue that is capable of providing e the desired representational functions. It has the potential for adaptation to individual users, broadening the reach of science programmes.

Our joint UCL/BBC project is developing a concept mapping second screen as a prototype for the Supernovae programme. The prototype will be used to demonstrate that a concept mapping second screen is able to foster learning from watching science on the television; is at an appropriate level of abstraction at which users will wish to interact, and; offers possibilities for augmenting different users differently so that the broadest audience can be both engaged and informed by science programmes.

REFERENCES

1. BBC, *The Frozen Planet Dual Screen.* 2012.
2. BBC, *The Seven Ages of Starlight.* 2012.
3. Cañas, A. J., Hill, G., Carff, R., Suri, N., Lott, J., Eskridge, T., Gómez, G., Arroyo, M. and Carvajal, R. CmapTools: A knowledge modeling and sharing environment. *Proc. first international conference on concept mapping*. 2004.
4. Cesar, P., Bulterman, D.C., and Jansen, A. Usages of the secondary screen in an interactive television environment: Control, enrich, share, and transfer television content, in *Changing television environments*. 2008, Springer. p. 168-177.
5. Davies, M., Concept mapping, mind mapping and argument mapping: what are the differences and do they matter? *Higher education*, 2011. **62**(3): p. 279-301.
6. De Cheveigné, S. and Véron, E. Science on TV: forms and reception of science programmes on French television. *Public Understanding of Science*, 5(3), 1996.
7. Hay, D.B., H. Wells, and I.M. Kinchin, Quantitative and qualitative measures of student learning at university level. *Higher Education*, 2008. **56**(2): p. 221-239.
8. Klein, J., Freeman, J., Harding, D., & Teffahi A. *Assessing the impact of second screen,* 2014. Ofcom.
9. Kornilakis, H., Grigoriadou, M., Papanikolaou, K. and Gouli E. Using WordNet to support interactive concept map construction. *Proceedings* of ICALT '04. IEEE.
10. Moulding, J., *BBC Reveals Companion Screen App Strategy*. Videonet, 2012.
11. Murray, J., et al. Story-map: iPad companion for long form TV narratives. *Proceedings of EuroiTV '12*. ACM.
12. Novak, J.D., Concept mapping: A useful tool for science education. *Journal of Research in Science Teaching*, 1990. **27**(10): p. 937-949.
13. Schäfer, M.S., Taking stock: A meta-analysis of studies on the media's coverage of science. *Public Understanding of Science*, 2012. **21**(6): p. 650-663.
14. Steyvers, M. and Tenenbaum, J.B. The Large-Scale Structure of Semantic Networks: Statistical Analyses and a Model of Semantic Growth. *Cognitive science*, 2005. 29(1): p.41-78.

Towards Multimodal Consumption of Georeferenced Mobile Video Using Shape and Speed

Sérgio Serra
LaSIGE, Faculdade de Ciências
Universidade de Lisboa
1749-016 Lisboa, Portugal
sergioserra99@gmail.com
+351217500087

Teresa Chambel
LaSIGE, Faculdade de Ciências
Universidade de Lisboa
1749-016 Lisboa, Portugal
tc@di.fc.ul.pt
+351217500087

ABSTRACT

An increasing amount of digital video is accessed, captured, and uploaded to the Web everyday, from different platforms and devices, that increasingly can georeference the information they capture and access, allowing to enrich their contextualization. But video search has been limited to keywords, or a set of parameters, providing limited support for temporal and spatial dimensions. We propose novel ways to search and access georeferenced videos, where these dimensions are of central importance, especially by video trajectories shape and speed, using a multimodal interactive mobile interface, involving gestures and movement, with the potential for more natural interactions, increased engagement, sense of presence and immersion.

The preliminary evaluation based on low-fidelity prototypes and encouraging users participation in the design, had positive results. Users found most features quite satisfactory, even fun, and easy to use. Different options and modalities were found interesting and adequate for different use scenarios that could be identified and suggested, and some concerns and challenges were identified to be taken into account in the next design and development phases, towards more flexible and effective interactive content consumption, through more natural interaction with mobile devices on their own or as second screens.

Author Keywords

Georeferenced videos; Gestures; Multimodal; Search; Browsing; Consumption; Space; Speed; Shape; Movement; Trajectories; 360º; Mobile; Second Screen

ACM Classification Keywords

H.5.1 [Information Interfaces and Presentation (I.7)]: Multimedia Information Systems – *video, hypertext navigation and maps*; H.5.2 [Information Interfaces and Presentation (I.7)]: User Interfaces – *interaction styles, evaluation*;

INTRODUCTION

Video is becoming a pervasive medium, widely captured, shared and accessed from different platforms and devices. And increasingly videos can be georeferenced, allowing to enrich their contextualization. A large amount of digital video is being uploaded everyday to the Web and is available to search and watch. However the current and most used mechanisms to browse and find videos are keywords and a limited set of parameters such as: keywords, duration, video quality, ignoring the temporal and spatial dimensions. Video has an enormous potential for immersion and mobile devices allow to access information while 'immersed' in reality anywhere. With the proliferation of devices like: smartphones, tablets and more recently wearables, we could take advantage of the multimodal sensors available, to create new ways to find and navigate georeferenced videos through time and space, using more natural interfaces, involving gestures and movement shape and speed, with the potential for increased engagement, sense of presence and immersion when accessing the videos.

In this paper, we describe our work in this direction. The next section presents the background provided by previous work and the vision for the new directions explored in this paper, followed by a section that highlights main challenges and opportunities, and presents most relevant related work. Next, the conceptual model and design options are presented for the multimodal georeferenced mobile video access in space and time, demonstrated in the prototypes and evaluated in the following section. A preliminary user evaluation was conducted with low fidelity prototypes, to find out about perceived usability and acceptance, focusing on usefulness, satisfaction and ease of use, and encouraging users participation in the design [12]. Finally, the paper ends with conclusions and perspectives for future work, also reflecting on questions relevant to effective interactive content consumption.

BACKGROUND AND VISION

This work builds on previous work done in the context of Sight Surfers [6], an interactive web application for sharing, visualizing and navigating georeferenced 360º interactive videos, as hypervideos, including city tours or more extreme activities like kart racing. These can be experienced in increased immersion and isolation, or synchronized with a map while being played. Sight Surfers supports several

2nd International Workshop on Interactive Content Consumption at TVX'14, June 25, 2014, Newcastle, UK.

mechanisms for navigation and orientation. Users can see the location and trajectory of the video in the map and navigate through crossing trajectories, possibly shot by other users, either in the map or as hyperlinks in the current video, and link to movie scenes that take place in that location. Windy Sight Surfers extended the previous system, to run on mobile devices and to empower users in their immersive video experiences [9,10]. It uses geographical and meteorological metadata, sensors and actuators, for increased immersion in terms of video viewing and sensing (visual, auditory and touch), for a more realistic feeling of movement and speed. It can be used on its own or as a second screen, having the video playing in a wide screen free of extraneous information and the mobile providing additional navigation and orientation features, e.g. using a map, for a more immersive experience.

Users can view around the 360º video by panning the device around, as if holding a window to the surrounding immersive video it is displaying, or to use it as a "wheel" - a second screen that can be rotated to move around the video in the wider TV screen. It may also provide an increased sense of speed and orientation when watching the videos through 3D and a wind interface. But videos are searched mainly by keywords, possibly filtered by regions in the map, time they were shot, duration, and broad categories ranging from fast to slow, in text and check-box based classical interfaces.

To provide a more complete support for the additional spatio-temporal dimension in georeferenced videos, and keeping the purpose of increasing immersion, aligned with the augmented sensorial experience, we want to create richer mechanisms for interactive search, visualization and navigation in more natural modes of interaction. Videos could be searched and accessed by their location as a place (e.g. New York, or the user's current position), as is often possible, but also based on their trajectories by choosing the actual streets, or even by the shape of the trajectories regardless of the actual streets (e.g. motocross or ski), by time (when were shot, and their duration) and by speed. Users could use touch interfaces for this, or use the mobility of their devices or their own movement while walking, running or traveling, on their own or as 2nd screens, to imprint or capture movement shape and speed, in possibly more natural and immersive ways.

RELATED WORK, CHALLENGES AND OPPORTUNITIES

Challenges for this work include providing users with an adequate interactive interface capable of capturing and expressing the temporal and spatial dimensions, allowing to represent speed and trajectories, and at the same time offering an intuitive, simple, effective and natural way to search for, and present resulting videos and navigate them in a mobile environment. It is both a challenge and an opportunity because users are not used to searching and navigating in these dimensions, but technology is allowing to capture movement in mobile devices in ways that hold the potential to support more natural interactions involving shape and speed towards more immersive experiences.

Most video libraries and websites like YouTube or Vimeo are based on keywords and have at most a very limited support to access video based on spatial and temporal dimensions. Rego et al. [11] developed VideoLIB, a digital library that enhances video retrieval by using spatial and temporal operators, based on Dublin Core and MPEG-7 metadata standards. Search criteria include action (what), person (who), time (when) and place (where), and use operators like before, during and after to define time intervals. This allows to make searches like "retrieve Madonnas´s video clips which were produced outside the USA during 1990's". It uses a form and text based interface, without the use of maps, and videos are considered as a whole - trajectories and speed are not taken into account.

There are some approaches to search and browse videos, and mainly photos, using maps. Google Street View is a 360º photo viewer using a spherical image projection and geolocalization, but it does not provide video, nor user generated and alternative views of the places. Panoramio (.com) is a georeferenced photo sharing website accessed as a layer in Google Earth and Google Maps. Users can do text-based search or navigate in the maps, and view photos taken by other users, based on location. The photos are presented along with a map that highlights their location, both as a collection resulting from a query or one by one. There are filters to highlight most popular, recent, famous places and indoor, both on a separate tab with the filtered photos, and by enlarging these photos among those shown on the map. Finsterwald et al. [2] developed The Movie Mashup Application (MOMA), as a public web map-based service for searching movies based on location, combining geotagged resources and text processing, mashing up information from DBpedia, GeoNames and Wikipedia synopses. Through its GUI, it allows to search and browse a data set of movies by director, location in text, by polygonal areas in the map, from locations extracted from movie titles, to compare query distributions and, using a mobile device version, allows to query for movies whose action took place around the user's current location. Although maps are a natural way to represent georeferenced information, and video often involves a trajectory, most solutions only allow users to post or access videos based on a single GPS location (usually the initial position). Seo et al. [14] present user-generated videos that relate to geographic areas in a map interface. They focus on the automatic selection of keyframes to represent the videos, and the determination of the location to place them on the maps. So they emphasize hotspots that are shot in the videos in front of the shooting spot, and not so much on their trajectories.

Concerning spatial and haptic interactive search in a mobile environment, in the last years we noticed a growing popularity of gesture interfaces and second screen applications. Lei & Coulton [3] implemented a gesture controlled application that act as a wand, using mobile sensors. It allows both proximity and remote search of points of interest (POIs) based on the orientation of the wand, as an interactive spatial 'Flashlight', and the possibility of users to create additional content for a particular POI as photographs tagged with POI's location and the direction from which the photograph was taken. Photographs can then be filtered based on a desired viewing angle in a real world environment. Premraj et al. [8] presented iWalk, a tool that allows multi-media exploration of geo-tagged data through movement, to move through the digital space of a collection, and gesture, for direct data manipulation (e.g. select, go to next, zoom). They experimented with geo-tagged photographs and sound collections, and a non geo-tagged museum collection, where the user defined a mapping between digital and physical spaces. Their approach makes use of computer vision algorithms computed on standard commercial camera inputs and is able to operate in real time.

Mobile devices may also act as second screens [1] to complement and interact with larger screens like TV or even public displays. MobiToss, an application created by Scheible et al [13], allows for mobile multimedia art sharing and creation. By using a mobile device with built-in accelerometer sensors, users can take a photo or video and "throw" it onto a large public display, with a gesture, for viewing and manipulation, through tilting. The users-created clips are augmented by the system with items like music or brand names and sent back to their phones as personal artifacts of the event. The preliminary user evaluation showed that capturing and throwing mobile content onto a large screen and manipulating it with gesture control into an art piece was perceived as an intuitive and fun activity. They enjoyed and engaged in the experience and appreciated getting something out of it, especially something artistic. But it requires improvements, by applying a more balanced set of video effects, adding group interaction and a more intuitive UI, to accommodate different movements users do to throw, and increase the perception of what is going on. This work explores natural gestures with a mobile device to manipulate photos or videos as a second screen, but in doing so, it does not explore spatial and temporal dimensions in videos. And none of the related work found addresses speed and trajectories in video as we propose to do.

VIDEO ACCESS BY SPACE, TIME AND… SPEED

Space and time dimensions are taken into account primarily in videos' locations, trajectories shapes and speed. To explore the interactive interfaces and user experience with these dimensions in georeferenced videos, a set of sketches and low-fidelity prototypes with different variations were designed for different use scenarios. The high-fidelity prototypes will explore the use of sensors and location services. The early low-fi prototyping and evaluation - using paper based screens on top of a real smart phone, users gestures and imagination - allow to test different design alternatives to tackle each goal in a faster and more efficient way, addressing challenges, sorting out and refining options, since an early stage. Next, we present the design rationale behind the main options for searches and accesses based on trajectory speed (Fig.1) and shape (Fig.2) in different modalities.

Through Touch – with finger

This is the more conventional interface, that allows to draw the query shapes by touching the screen (Fig.2b), or even on a touch pad of a laptop or desktop. The speed of this drawing may also be captured to query by speed only, or by both shape and speed. This kind of interaction may be more familiar and provide for better accuracy than the following, especially when the user has a hand free for this interaction

Through Gesture – with mobile

This modality can be used by moving the mobile in a gesture, to draw a shape (Fig.2a) or demonstrate a speed level (Fig.1a). This can be done with the hand that is holding the device, even if the other one is not free, and has the potential for a more natural or immersive modality to select the desired videos, to watch on the mobile or on a wider screen like a TV. In this context, viewers are used to interacting with a control in one hand, and keeping focused on the video on the wide screen.

Through Traveling – on the move

When on the move, users are often travelling by car, train, subway, planes or even walking or running. In addition or as an alternative to use your current location to access videos shot in the same location, it can be interesting to take the chance, especially when not driving, to watch videos that were shot at a similar speed, and be able to enjoy the viewing experience in a more immersive way, by matching the speed of what you are seeing with the speed that you are feeling or experiencing in reality. This might have a special impact in high speed videos in more extreme activities. As in the previous cases, both speed and trajectory shape can be captured this way (Fig.1b). Speed and even location might have more potential for immersion in the immediate video viewing, to get related videos on the spot. But capturing a trajectory can also be interesting, to search for other videos that have similar paths even if in different places, e.g. another similar Kart race elsewhere in the world. While gestures could rely on sensors, travel features would rely on location services like GPS.

Where - on the Map or Current Location

Since Sight Surfers videos are georeferenced, search can be made dependent on their location. Users can use a map to select locations (Fig.2b), or the current location can be captured, for the queries (besides the possibility to specify a set of places, like cities, as in more traditional interfaces).

Anywhere

Videos may also be searched independent of their location. In this way, a map is not used and only speed (Fig.1a) and shape are drawn on the screen or in the air (Fig.2a), or captured on travel (Fig.1b), without a geographical reference.

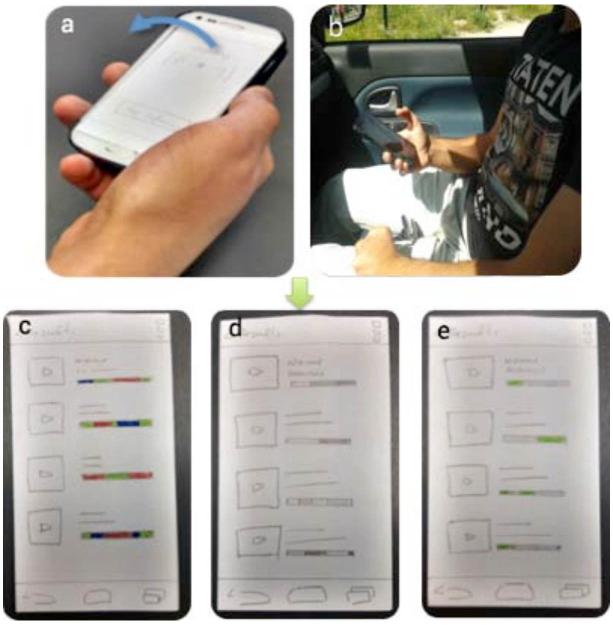


Figure 1. Search by Speed, through: a) gesture; b) traveling speed on a car; Results in video List with: c) Colored timelines with 3 colors; d) Gray-scale timelines; e) Color Highlight timelines with green color for the searched speed.

Results in Maps or Lists

The resulting videos can be presented as trajectories on a map (Fig.2d), where each trajectory can also be seen as the video timeline synchronized with the video as in Sight Surfers [6]. And this is the default when search is based on a location. But results may also be presented independent of their location, e.g. in a list, where the speed (Fig. 1c-e) and or the shape (Fig.2c) can be emphasized in each video timeline. Also note that users could switch between map and list views and select what to show, for the same results.

Search results are presented to the user in different design alternatives, each one offering different visual cues and information about the content retrieved, in terms of shape and speed, both in maps or lists.

Speed Awareness

Speed can vary along a video, so the results would present first the videos that keep the desired speed (with a tolerance) for longer, but still, it could be interesting to be aware of the segments where the speed is as queried for. Fig.1c-e shows three alternative designs for presenting speed in the video *timelines* in: 1) Color: green for the searched speed, red for faster and blue for slower; 2) Gray-Scale: mid tone for searched speed, darker for faster and lighter for slower; and 3) Color Highlight: green for searched speed and two gray tones for faster an slower as in 2), allowing for higher contrast of the searched for speed.

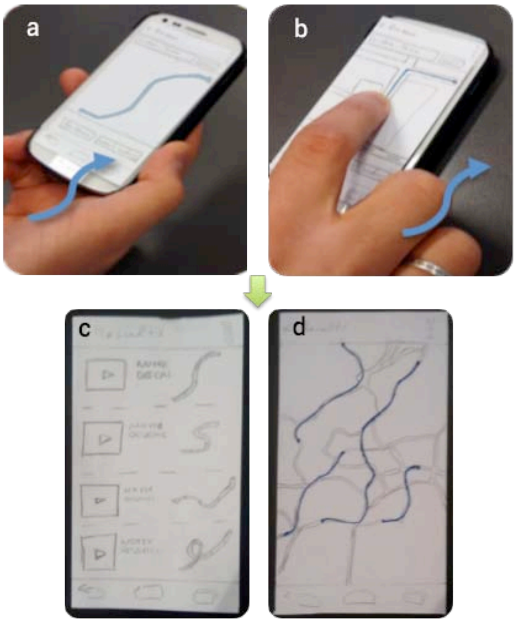


Figure 2. Search by Shape, through: a) gesture; b) geo-referenced shape with touch; c) Results on a List with video shaped timelines; d) Results on a Map showing videos trajectories.

Shape Awareness

Shape is shown by default when on a map, but can also be presented in the list view, as in Fig.2c where each video timeline takes the shape of the corresponding trajectory. Whereas speed awareness is optional in all the timelines: map, and list with or without shape.

PRELIMINARY USER EVALUATION

We conducted a user study to evaluate the features designed and to investigate about preferred alternatives and users' perception about usability and user experience, and their application in real use scenarios.

Method

We performed a task-oriented evaluation based mainly on Observation and semi-structured Interviews, after explaining the purpose of the evaluation and the concept behind the Sight Surfers application context and the new features being evaluated in the low-fidelity prototypes, using a Wizard of Oz approach for the interaction feedback. At the end of each task and at the end as overall, users provided a 1-5 USE (Usefulness, Satisfaction, and Ease of use) rating [4] about the tested interactive features, and were encouraged to make comments and suggestions, that at the current stage had the potential for contributions in a participatory design perspective [12].

Participants

There were 10 participants aged 21-52 (24 on average, 4 F, 6 M). All users had at least finished high school, 3 from computer science, the rest a mix of backgrounds, all had a smart phone used on a daily basis to access info, and 9 often search for and watch videos but mainly on PCs, sometimes on tablets and seldom on mobiles.

Results

Main results are summarized by mean values for USE and most significant comments, for each of the categories of features. Std deviation was 0.5-1.0 (in a 1-5 scale) reflecting some different opinions that are commented.

Search by speed through gestures with the mobile was considered quite fun and very easy to use, and found useful by some (U:2.6;S:3.1;E:4.1), e.g. *"I can imagine being a nice thing to have when you practice some sort of motorsport or if you like to watch sports like Formula 1, guys that like speed would probably like this",* but others had doubts about its usefulness in real life, on a daily basis. Some concerns about the device precision to capture the wanted speed were also raised, especially by the computer science students. Travel speed was considered more useful in a real world scenario than moving the phone to capture speed (3.2;3.2;4.0). *"It is more useful for example for someone that practices motorsports", "I prefer it, I can use it to know the speed I'm going and look for videos with it".*

Search by shape with touch was considered quite fun and easy to use, but users were not sure about its usefulness (U:2.5;S:3.1;E:3.8): *"It could be fun sometimes, but normally we want to search for things".* On the other hand, participants appreciated having the ability to georeference the shape, finding several use scenarios where it could be used. *"I think georeferenced video search is much more useful, I can use it to see videos from a particular street, city, whatever", "It's better than normal shapes, I'm an athlete and I can use it to find and see running tracks".* Some users pointed out that it is not easy to draw the trajectory *"In Google maps I can click the points I want and it draws the trajectory automatically".* Searching for videos with a specific shape by using the phone to draw was found easier for free shapes (3.6;3.3;4.2), and not so much for geo-referenced shapes (2.8;2.4;3.0), but less precise. Users are not used to using the phone for this kind of action, preferring the finger to do it. *"it is not practicable, imagine walking and doing it on the street", "ok with a phone, but maybe not with a larger device like a tablet".*

In general, users preferred the more familiar way of showing the videos in a **list view** (U:4.3;S:4.0;E:4.6), by being easy and simple, *"less fancy but does the work".* But some fancied the **map** idea better (3.2;3.2;3.2) for the additional information, and most said it was useful when looking for georeferenced shapes, videos in specific places, to be aware of the video locations and trajectories length. *"As a geographical engineer, this could be useful for my work".* Main concerns referred to awareness of the amount of videos retrieved and the representation in the presence of a huge amount. Filtering of results [6] was not in the scope of this test, but aligned with their concerns. A user suggested to have a mix of list and map where one could spot the videos on a map while hovering through a list of the selected ones, and on the other way around, hovering a trajectory could show the info that is presented in a list entry (with speed, duration, video image, etc.) in a popup with more detailed information.

Users found the **timelines** useful, satisfactory and easy to use for shape and speed awareness. Regarding speed and the different designs: *"are a nice idea to have because they easily indicate the speed of the video",* although a couple of users mentioned they would only possibly needed to identify the searched speed and not distinguish higher and lower speeds. Most users preferred the color highlight (U:4.0; S:4.0; E:4.6) in the search speed with the other speeds made less noticeable in gray for being easier to use, *"It's less confusing and very simple to use, easier than the 3 colors"*; and then the color version (4.3;4.1;3.9), which they found useful and satisfactory although more difficult to use. The gray version was found more difficult to use and less useful, satisfactory and even fun (1.9;1.7;2.4).

Time and Space Revisited

Space was addressed in georeferencing and represented in maps and in trajectories' shapes. Time is inherent to the video and represented in the timelines, and combined with space can be represented as trajectories in maps that can be presented in synchrony with the videos [6] and was now also addressed in speed. For its more natural mapping, the interactions that involved space, and especially shape and speed, received more attention in the conception of natural interfaces so far.

In the near future, we intend to explore further the temporal dimension in search and navigation, both inside each video and among videos that were shot in different periods in time. Although we have some ideas about the design, we wanted to learn from the users about their visions for interactions involving the temporal dimension in a more participatory way. Almost all participants immediately associated time with a timeline, like the ones we already have in the prototype in the different designs. A similar concept could be extended to refer to the time when the videos where shot, allowing to select temporal moments or intervals, and the possibility to enter specific dates in a text field was also mentioned. They acknowledged the importance of this dimension and the way it was already addressed. It was not so easy for them to move from what was already familiar, although they were curious and open about the possibility of also using different modalities.

CONCLUSIONS AND FUTURE WORK

We presented the motivation and design options for georeferenced mobile video access in space and time, developed in a context of user generated content, with a special focus on shape and speed in different modalities based on touch, gestures and movement. The preliminary evaluation based on low-fidelity prototypes had encouraging results. Users found most features quite satisfactory, even fun, and easy to use, and different options and modalities were found interesting and adequate for different use scenarios that could be identified. But although some users found these interesting uses, others were skeptical about the usefulness of using trajectories and speed for video access in real life, something they are not used to having. There were also some concerns about sensor accuracy in gesture modalities.

It is important to highlight that evaluating this type of features on low-fidelity prototypes does not offer the same experience than a high fidelity application running on the smartphone, even more when using sensors and viewing videos – both providing dynamic information - are central aspects in the system. Also some users tended to more familiar ground and conservative options, in this first contact with these features, especially the ones with less technical background, or not used to accessing georeferenced media. Still, the evaluations, comments and suggestions allowed to identify upfront main strengths and concerns, and aspects to improve and reinforce in the design, implementation and reevaluation as a high fidelity prototype, that is already in progress as the next iteration. As innovative interaction modalities, based on sensors and location services, there are technological challenges in terms of accuracy and smooth integration of modalities that are being taken into account for the effectiveness of the designed interactions.

The temporal dimension in search and navigation will also be further explored, taking the feedback received into account, along with the enrichment of navigation when viewing the video. The adoption of these more natural interactions also in this context, and building on our previous work on immersive video [9,10] has the potential to increase the sense of presence and immersion when navigating around, crossing trajectories, changing viewing speed, or even throwing the video [1] to view on a wider screen like a TV, while possibly keeping the mobile as a second screen, or capturing the video playing on the TV to the mobile, to go on watching it on the move.

Closing remarks: The interactive forms of video access when the user is more active and often in a more reflective cognitive mode [5] are interweaved with moments of more passive consumption when the user views the video in a more experiential cognitive mode. We believe that more natural and immersive forms of viewing and interacting with video influence the quality of the user experience, mainly in experiential modes, and that a good design can contribute to foster and support the reflective modes in complementing and harmonious ways, so important in learning [7]. We are following previous research work [7,6,9,10] with the goal to contribute in this direction towards richer, more effective and satisfactory consumption of video-based media in different scenarios.

ACKNOWLEDGMENTS

This work is partially supported by FCT through funding of LaSIGE Strategic Project, ref. PEst-OE/EEI/UI0408/2014 and the ImTV research project ref. UTA-Est/MAI/0010/2009.

REFERENCES

1. Courtois, C., D'heer, E. Second screen applications and tablet users: constellation, awareness, experience, and interest. Proc. EuroiTV'12, ACM Press (2012),153-156.
2. Finsterwald, J. M., Grefenstette, G., Law-To, J., Bouchard, H., and Mezaour, A.D. The Movie Mashup Application MoMa: Geolocalizing and Finding Movies. Proc. GeoMM'12, ACM Press (2012), 15-18.
3. Lei, Z., and Coulton, P. A. Mobile Geo-wand Enabling Gesture Based POI Search an User Generated Directional POI Photography. Proc. ACE'09, the Int. Conf. on Advances in Computer Entertainment Technology, ACM Press (2009), 392-395.
4. Lund, A. M. Measuring usability with the USE questionnaire. Usability and User Experience, 8(2), (2001).
5. Norman, D. Things that Make us Smart. Addison Wesley Publishing Company (1993).
6. Noronha, G., Álvares, C., and Chambel, T. Sight Surfers: 360º Videos and Maps Navigation. Proc. GeoMM'12, ACM Press (2012), 19-22.
7. Prata, A., and Chambel, T. Going Beyond iTV: Designing Flexible Video-Based Crossmedia Interactive Services as Informal Learning Contexts. Proc. EuroITV'2011, ACM Press (2011), 65-74.
8. Premraj, V., Schedel, M., and Berg, T.L. iWalk, A Tool for Interacting with Geo-Located Data Through Movement and Gesture. Proc. ACM MM'10, ACM Press (2010), 1059-1062.
9. Ramalho, J., and Chambel, T. "Immersive 360º Mobile Video with an Emotional Perspective". Proc. ImmersiveMe'2013, ACM Press (2013), 35-40.
10. Ramalho, J., and Chambel, T. "Windy Sight Surfers: Sensing and Awareness of 360º Immersive Videos on the Move". Proc. EuroiTV'2013, ACM Press (2013), 107-115.
11. Rego, A., Baptista, C., Silva, E., Schiel, U., and Figueirêdo, H. VideoLib: a Video Digital Library with Support to Spatial and Temporal Dimensions. Proc. SAC'07, ACM Press (2007), 1074-1078.
12. Sanders, E. From User-Centered to Participatory Design Approaches. In Design and the Social Sciences. J. Frascara (Ed), Taylor & Francis Books Limited (2002).
13. Scheible, J., Ojala, T., and Coulton, P. MobiToss: A novel gesture based interface for creating and sharing mobile multimedia art on large public displays. Proc. ACM MM'08, ACM Press (2008), 957-960.
14. Seo, B., Hao, J., and Wang, G. Sensor-rich Video Exploration on a Map Interface. Proc. ACM MM'11, ACM Press (2011), 1013-1016.

Organizers

Britta Meixner is a researcher at the Passau University. She received a diploma in computer science and a state examination for lectureship at secondary schools from the University of Passau, Germany, in 2008. Currently, she is working towards a PhD degree in computer science at the Faculty of Computer Science and Mathematics of the Passau University. There, she is conducting research and development in the area of hypervideo. Thereby, she focuses on creating easy to use authoring tools and players and on download and cache management to provide a better viewer experience. Further she is interested in hypervideo on mobile devices, collaborative hypervideo creation, and decision rules in hypervideo. Britta is a member of the BMBF research project “Mirkul” that investigates application scenarios of interactive nonlinear video. Britta is a reviewer for the Multimedia Tools and Applications Journal (Springer) and was a member of the program committee of the 1st International Workshop on Interactive Content Consumption at EuroITV 2013.

Katrin Tonndorf is a researcher at Passau University. She received a magister degree in media studies from the Technical University Braunschweig and the Braunschweig University of Arts in 2010. Currently, she is working towards a PhD degree in communication studies at the Faculty of Arts and Humanities at the Passau University. She is conducting research in the area of online and social media communication practices. Furthermore she is interested in the use of interactive audiovisual content for learning und support purposes. Katrin is also a member of the BMBF research project “Mirkul”.

Rene Kaiser is a key researcher for JOANNEUM RESEARCH and has been involved in a number of European projects dealing with automation of content production such as NM2, Aposdle, TA2 and Vconect. His research focus is on Virtual Director software, on automating shot selection through cinematographic behavior models. Further he is interested in automating non-linear video production, enabling the user to interactively influence the narrative path while watching. Rene was responsible for the organization of the Interactive and Immersive Entertainment and Communication Special Session at MMM’12. He is part of a group hosting the annual PhD cooperation workshop at the i-KNOW and i-SEMANTICS conference, active member of STCSN, and has been organizing the Barcamp Graz, a yearly 3-day unconference which is an interactive and open discussion format. At EuroITV 2013, Rene was co-organizing the first edition of WSICC.

Joscha Jaeger is a research assistant at Merz Akademie Stuttgart and founder of filmicweb - Hypervideo Interface Design. His research covers web-based hypervideo technology, time-based interaction and semantic video search interfaces. Joscha has a strong focus on film as information architecture, collaborative editing systems for non-linear film and user-driven annotation systems. He is interested in finding new ways of distributed interaction with open video technologies and interfaces on the web.

The workshop was supported by the European Commission under the contract FP7-287760, “Vconect - Video Communications for Networked Communities” and by the German Federal
Ministry of Education and Research (BMBF) under contract 03V0633